

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2004, 19:19:41 / Search time 17 seconds

(without alignments)
980.144 Million cell updates/sec

Title: US-09-936-456-2

Perfect score: 320

Sequence: 1 MTLWNGVLPFYQPRHAGF.....GDPHQKQALPDKCITNTL 320

Scoring table: OLIGO
Gapop 60.0, Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size: 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	2.5	615	1 MUTL_ECO57	Q8XDN4 escherichia
2	8	2.5	615	1 MUTL_ECOL6	Q8EAK9 escherichia
3	8	2.5	615	1 MUTL_ECOL1	P23367 escherichia
4	8	2.5	617	1 MUTL_PASMT	P57886 pasteurella
5	8	2.5	618	1 MUTL_SALTI	Q8Z187 salmoneilla
6	8	2.5	618	1 MUTL_SALTY	P14161 salmoneilla
7	8	2.5	624	1 MUTL_CHUTE	Q8KXK3 chlorobium
8	8	2.5	629	1 MUTL_HAEIN	P44494 haemophilus
9	8	2.5	632	1 MUTL_PSEPK	Q884d1 pseudomonas
10	8	2.5	633	1 MUTL_PSEPE	Q9N118 pseudomonas
11	8	2.5	635	1 MUTL_YARPE	Q8Z144 yersinia pe
12	8	2.5	645	1 MUTL_PESPM	Q874J2 pseudomonas
13	8	2.5	645	1 MUTL_STRAP	Q8CEP9 staphylococ
14	8	2.5	653	1 MUTL_VIBCH	Q9KX13 vibrio chol
15	8	2.5	664	1 MUTL_VIBVU	Q8DCV0 vibrio vuln
16	8	2.5	669	1 MUTL_STRAU	Q994H7 staphylococ
17	8	2.5	669	1 MUTL_STRAU	Q93105 staphylococ
18	8	2.5	669	1 MUTL_STRAU	Q8KXK9 staphylococ
19	8	2.5	669	1 MUTL_VIBPA	Q87105 vibrio para
20	8	2.5	49	1 COXA_PARDE	P77921 paracoccus
21	7	2.2	116	1 CHPB_ECOLI	P33647 escherichia
22	7	2.2	116	1 CN4D_MOUSE	Q01063 mus musculu
23	7	2.2	131	1 PRDC_PROVU	P20923 proteus vul
24	7	2.2	181	1 IGFP2_PIG	P23695 sus scrofa
25	7	2.2	243	1 YGBR_BACSU	P42250 bacillus su
26	7	2.2	262	1 COBS_SHRON	Q81174 shewanella
27	7	2.2	266	1 FLIY_ECOLI	P39117 escherichia
28	7	2.2	277	1 EFB_STRAU	P09332 staphylococ
29	7	2.2	281	1 BPHB_COMTE	Q46381 c cis-2,3-d
30	7	2.2	288	1 YGBH_ECOLI	Q46845 escherichia
31	7	2.2	289	1 ATP6_PROMO	P21903 propionigen
32	7	2.2	299	1 T2RI_HUMAN	Q9NW77 homo sapien
33	7	2.2	308	1 KHSE_CORGL	P07128 corynebacte

ALIGNMENTS

```

RESULT 1
MUTL_ECO57          STANDARD;          PRT;          615 AA.
ID      MUTL_ECO57
AC      Q8XDN4;
DT      28-FEB-2003 (Rel. 41, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      DNA mismatch repair protein mult.
GN      MUTL_OR_Z5777 OR EGS5146.
OS      Escherichia coli O157:H7.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OX      Enterobacteriaceae; Escherichia.
ON      NCBI_Taxid=83334;
OR      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=O157:H7 / BDJ933 / ATCC 700927;
RX      MEDLINE=21074935; PubMed=11206551;
RA      Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA      Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA      Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller J.,
RA      Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocitis K.,
RA      Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA      Welch R.A., Blattner F.R.,
RT      "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL      Nature 409:529-533 (2001).
RM      [2]
RN      SEQUENCE FROM N.A.
RC      STRAIN=O157:H7 / RIMD 0509952;
RX      MEDLINE=21156231; PubMed=11258796;
RA      Hayashi T., Makino K., Kurokawa K., Ishii K., Yokoyama K.,
RA      Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA      Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
RA      Kuhara S., Shiba T., Hattori M., Shinagawa H.,
RT      "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT      O157:H7 and genomic comparison with a laboratory strain K-12.";
RL      DNA Res. 8:11-22(2001)
-!- FUNCTION: This protein is involved in the repair of mismatches in
-!- DNA. It is required for dam-dependent methyl-directed DNA mismatch
-!- repair. May act as a "molecular matchmaker", a protein that
-!- promotes the formation of a stable complex between two or more
-!- DNA-binding proteins in an ATP-dependent manner without itself
-!- being part of a final effector complex (By similarity).
-!- SIMILARITY: Belongs to the DNA mismatch repair mult/hexB family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AE005650; AAC59366.1; -
EMBL; AP002568; BAB38569.1; -
PIR; B66113; B66113.
PIR; B91272; B91272.

```

DR HAMAP; MF_00149; -, 1.
 DR InterPro; IPR003594; ATPbind_ATPase.
 DR InterPro; IPR002099; DNA_mis_repair.
 DR Pfam; PF01119; DNA_mis_repair; 1.
 DR Pfam; PF02518; HATPase_C; 1.
 DR SMART; SM00387; HATPase_C; 1.
 DR TIGRfam; TIGR00585; mutL; 1.
 DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
 KW DNA repair; Complete proteome.
 SQ SEQUENCE 615 AA; 67921 MW; 1B5F0177AB11EB9 CRC64;

Query Match 2.5%; Score 8; DB 1; Length 615;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
 DB 98 ALASISSV 105

RESULT 2

MTL_EC0L6 STANDARD; PRT; 615 AA.
 AC 08FAK3;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE DNA mismatch repair protein mutL.
 GN MUTL OR C5254.
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=217992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=06.H1 / CFT073 / ATCC 700928;
 RX MEDLINE=22386234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesech P.,
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic Escherichia coli.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).

CC -1- FUNCTION: This protein is involved in the repair of mismatches in
 CC DNA. It is required for dam-dependent methyl-directed DNA mismatch
 CC repair. May act as a "molecular matchmaker", a protein that
 CC promotes the formation of a stable complex between two or more
 CC DNA-binding proteins in an ATP-dependent manner without itself
 CC being part of a final effector complex (By similarity).
 CC -1- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sb.ch/announce/>
 CC or send an email to license@isb-sb.ch).

CC EMBL; AEO16771; XAN83676.1; ALT_INIT.
 DR HAMAP; MF_00149; -, 1.
 DR InterPro; IPR003594; ATPbind_ATPase.
 DR InterPro; IPR002099; DNA_mis_repair.
 DR Pfam; PF01119; DNA_mis_repair; 1.
 DR Pfam; PF02518; HATPase_C; 1.
 DR TIGRfam; TIGR00585; mutL; 1.
 DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
 KW DNA repair; Complete proteome.
 SQ SEQUENCE 615 AA; 67910 MW; 770FA42138081CBB CRC64;

Query Match 2.5%; Score 8; DB 1; Length 615;
 Best Local Similarity 100.0%; Pred. No. 12;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
 DB 98 ALASISSV 105

RESULT 3

MTL_EC0L1 STANDARD; PRT; 615 AA.
 ID MUTL_EC0L1
 AC P23367;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE DNA mismatch repair protein mutL.
 GN MUTL OR B4170.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KL2;
 RX MEDLINE=93279041; PubMed=1594459;
 RA Tsui H.-C., Mandavilli B., Winkler M.E.;
 RT "Nonconserved segment of the mutL protein from Escherichia coli K-12
 and Salmonella typhimurium.";
 RL Nucleic Acids Res. 20:2379-2379(1992).

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KL2 / MG1655;
 RX MEDLINE=9534362; PubMed=7610040;
 RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
 RA Blattner F.R.;
 RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
 RT region from 92.8 through 100 minutes.";
 RL Nucleic Acids Res. 23:2105-2119(1995).

RN [3]
 RP SEQUENCE OF 596-615 FROM N.A.
 RC STRAIN=KL2;
 RX MEDLINE=91154127; PubMed=1999389;
 RA Winkler M.E., Connolly D.M.;
 RT "Structure of Escherichia coli K-12 *mutL* and characterization of the
 RT mutator phenotype caused by *mutL* insertion mutations.";
 RL J. Bacteriol. 173:1711-1721(1991).

RN [4]
 RP SEQUENCE OF 1-23 FROM N.A.
 RC STRAIN=KL2;
 RX MEDLINE=94195106; PubMed=7511774;
 RA Tsui H.-C., Zhao G., Feng G., Leung H.-C., Winkler M.E.;
 RT "The *mutL* repair gene of Escherichia coli K-12 forms a superoperon
 RT with a gene encoding a new cell-wall amidase.";
 RL Mol. Microbiol. 11:189-202(1994).

RN [5]
 RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 1-349.
 RC MEDLINE=95043508; PubMed=9827806;
 RX Ban C., Yang W.;
 RA Ban C., Yang W.;
 RT "Crystal structure and ATPase activity of MutL: implications for DNA
 RT repair and mutagenesis.";
 RL Cell 95:541-552(1998).

RN [6]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-349.

RC STRAIN=KL2;
 RX MEDLINE=95213497; PubMed=10199405;
 RA Ban C., Junop M., Yang W.;
 RT "Transformation of MutL by ATP binding and hydrolysis: a switch in
 RT DNA mismatch repair.";
 RL Cell 97:85-97(1999).

CC -1- FUNCTION: THIS PROTEIN IS INVOLVED IN THE REPAIR OF MISMATCHES IN
 CC DNA. IT IS REQUIRED FOR DAM-DEPENDENT METHYL-DIRECTED DNA MISMATCH
 CC REPAIR. MAY ACT AS A "MOLECULAR MATCHMAKER", A PROTEIN THAT
 CC PROMOTES THE FORMATION OF A STABLE COMPLEX BETWEEN TWO OR MORE
 CC DNA-BINDING PROTEINS IN AN ATP-DEPENDENT MANNER WITHOUT ITSELF

BEING PART OF A FINAL EFFECTOR COMPLEX. THE ATPASE ACTIVITY OF
 MUTL IS STIMULATED BY DNA.
 -1- SIMILARITY: Belongs to the DNA mismatch repair mutl/hexb family.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 CC EMBL; Z11831; CAAT7850.1; -
 DR EMBL; U14003; AAA97066.1; -
 DR EMBL; AE000489; AAC77127.1; -
 DR EMBL; M63655; AAA24173.1; -
 DR EMBL; L19346; AAA20098.1; -
 DR PIR; PH0853; PH0853.
 DR PDB; 1B62; 28-APR-99.
 DR PDB; 1B63; 08-JUN-99.
 DR PDB; 1BKN; 11-MAY-99.
 DR EcoGene; EG11281; mutl.
 DR HAMAP; MF_00149; -; 1.
 DR InterPro; IPR003594; ATPbind_ATPase.
 DR InterPro; IPR002099; DNA_mis_repair.
 DR Pfam; PF01119; DNA_mis_repair; 1.
 DR Pfam; PF02518; HATPase_c; 1.
 DR SMART; SM00387; HATPase_c; 1.
 DR TIGRFAMs; TIGR00585; mutl; 1.
 DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
 KW DNA repair; ATP-binding; 3D-structure; Complete proteome.
 FT HELIX
 FT 8
 FT 19
 FT HELIX
 FT 23
 FT 36
 FT TURN
 FT 37
 FT 38
 FT STRAND
 FT 41
 FT 47
 FT HELIX
 FT 48
 FT 50
 FT TURN
 FT 51
 FT 51
 FT STRAND
 FT 53
 FT 58
 FT HELIX
 FT 65
 FT 67
 FT HELIX
 FT 68
 FT 72
 FT TURN
 FT 74
 FT 75
 FT HELIX
 FT 83
 FT 87
 FT TURN
 FT 88
 FT 103
 FT HELIX
 FT 98
 FT 103
 FT TURN
 FT 104
 FT 105
 FT STRAND
 FT 107
 FT 113
 FT TURN
 FT 115
 FT 116
 FT STRAND
 FT 120
 FT 126
 FT TURN
 FT 127
 FT 130
 FT STRAND
 FT 131
 FT 137
 FT STRAND
 FT 142
 FT 148
 FT TURN
 FT 150
 FT 153
 FT HELIX
 FT 155
 FT 159
 FT TURN
 FT 160
 FT 160
 FT HELIX
 FT 164
 FT 181
 FT TURN
 FT 183
 FT 184
 FT STRAND
 FT 186
 FT 191
 FT TURN
 FT 192
 FT 193
 FT STRAND
 FT 194
 FT 199
 FT TURN
 FT 204
 FT 205
 FT TURN
 FT 208
 FT 208
 FT HELIX
 FT 209
 FT 216
 FT STRAND
 FT 218
 FT 223
 FT TURN
 FT 224
 FT 231
 FT STRAND
 FT 232
 FT 233
 FT TURN
 FT 234
 FT 241
 FT STRAND
 FT 243
 FT 245
 FT HELIX
 FT 248
 FT 249
 FT TURN
 FT 250
 FT 252
 FT HELIX
 FT 254
 FT 258
 FT STRAND
 FT 259
 FT 260
 FT TURN
 FT 261
 FT 262
 FT STRAND
 FT 262
 FT 262
 FT HELIX
 FT 266
 FT 279

FT STRAND 288 293
 FT HELIX 286 298
 FT STRAND 289 300
 FT TURN 302 303
 FT TURN 305 306
 FT STRAND 311 312
 FT TURN 313 314
 FT HELIX 315 329
 FT TURN 330 330
 SQ SEQUENCE 615 AA; 67923 MW; C69D2735BF5FA165 CRC64;

Query Match 2.5%; Score 8; DB 1; Length 615;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 223 ALA51SSV 230
 Db 98 ALA51SSV 105

RESULT 4
 MUTL_PASMU STANDARD; PRT; 617 AA.
 ID MUTL_PASMU
 AC P57886;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE DNA mismatch repair protein mutl.
 GN MUTL OR PM0904.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Pasteurella.
 OC NCBI_taxid=747;
 OX NCBI [1]
 RN P
 RP SEQUENCE FROM N.A.
 RC STRAIN=Pm70;
 RX MEDLINE=21145866; PubMed=11248100;
 RA May B.J., Zhang O., Li L.L., Paustian M.L., Whitlam T.S., Kapur V.,
 RT "Complete genomic sequence of Pasteurella multocida Pm70.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
 CC -1- FUNCTION: This protein is involved in the repair of mismatches in
 CC DNA. It is required for dam-dependent methyl-directed DNA mismatch
 CC repair. May act as a "molecular matchmaker", a protein that
 CC promotes the formation of a stable complex between two or more
 CC DNA-binding proteins in an ATP-dependent manner without itself
 CC being part of a final effector complex (By similarity).
 CC -1- SIMILARITY: Belongs to the DNA mismatch repair mutl/hexb family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE006129; AAK02988.1; -
 DR HSSP; P23367; IBKN.
 DR HAMAP; MF_00149; -; 1.
 DR InterPro; IPR003594; ATPbind_ATPase.
 DR InterPro; IPR002099; DNA_mis_repair.
 DR Pfam; PF01119; DNA_mis_repair; 1.
 DR Pfam; PF02518; HATPase_c; 1.
 DR SMART; SM00387; HATPase_c; 1.
 DR TIGRFAMs; TIGR00585; mutl; 1.
 DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
 KW DNA repair; Complete proteome.
 SQ SEQUENCE 617 AA; 63798 MW; 8DC85F15695460AE CRC64;

Query Match 2.5%; Score 8; DB 1; Length 617;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISV 230
 DB 98 ALASISV 105

RESULT 5

MUTL_SALTY
 ID MUTL_SALTY STANDARD; PRT; 618 AA.
 AC Q82187;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE DNA mismatch repair protein mutL.
 DE MUTL OR STY4716 OR T4410.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaiha M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.W., Dowd L., White N., Farrar J.,
 RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
 RA Krogsh A., Larsen T.S., Leather S., Moule S., O'Goird P., Parry C.,
 RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrett B.G.;
 RT "Complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhi CT18.";
 RL Nature 413:848-852(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TY2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RA Deng W., Lou S.-R., Plunkett G., III, Mayhew G.F., Rose D.J.,
 RA Burdand V., Kodoyanni V., Schwartz D.C., Blattner F.R.,
 RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
 RT and CT18.";
 RL J. Bacteriol. 185:2330-2337(2003).
 CC -1- FUNCTION: This protein is involved in the repair of mismatches in
 CC DNA. It is required for dam-dependent methyl-directed DNA mismatch
 CC repair. May act as a "molecular matchmaker", a protein that
 CC promotes the formation of a stable complex between two or more
 CC DNA-binding proteins in an ATP-dependent manner without itself
 CC being part of a final effector complex (By similarity).
 CC -1- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AL627283; CAD06836.1; -;
 DR EMBL: AE016849; AAC071859.1; -;
 DR HAMAP: MF_00149; -; 1
 DR InterPro: IPR003594; ATPbind_ATPase.
 DR InterPro: IPR002099; DNA_mis_repair.
 DR Pfam: PF01119; DNA_mis_repair; 1.
 DR Pfam: PF02518; HATPase_C; 1.
 DR SMART: SM00387; HATPase_C; 1.
 DR TIGRPFAM: TIGR00585; mutL; 1.
 DR PROSITE: PS00058; DNA_MISMATCH_REPAIR_1; 1.
 DR DNA repair; Complete proteome.
 KW PROSITE: PS00058; DNA_MISMATCH_REPAIR_1; 1.
 SQ SEQUENCE 618 AA; 67829 MW; 18DD900435180F4 CRC64;

Query Match 2.5%; Score 8; DB 1; Length 618;
 Best Local Similarity 100.0%; Pred. No. 12;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 223 ALASISV 230
 DB 98 ALASISV 105

RESULT 6

MUTL_SALTY
 ID MUTL_SALTY STANDARD; PRT; 618 AA.
 AC P1461;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE DNA mismatch repair protein mutL.
 DE MUTL OR STM4359.
 GN Salmonella typhimurium.
 OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2;
 RX MEDLINE=90008765; PubMed=2676972;
 RA Markovitch J.A., McIntyre C.A., Walker G.C.;
 RT "Nucleotide sequence of the Salmonella typhimurium mutL gene required
 RT for mismatch repair: homology of MutL to HexB of Streptococcus
 RT pneumoniae and to PMS1 of the yeast Saccharomyces cerevisiae.";
 RL J. Bacteriol. 171:5325-5331(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewel N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2.";
 RL Nature 413:852-856(2001).
 CC -1- FUNCTION: This protein is involved in the repair of mismatches in
 CC DNA. It is required for dam-dependent methyl-directed DNA mismatch
 CC repair. May act as a "molecular matchmaker", a protein that
 CC promotes the formation of a stable complex between two or more
 CC DNA-binding proteins in an ATP-dependent manner without itself
 CC being part of a final effector complex.
 CC -1- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M29687; AAA27166.1; -;
 DR EMBL: AE008904; AAL21179.1; -;
 DR PIR: A33588; A33588.
 DR HSSP: P23367; 1B62.
 DR StyGene; SGI0236; mutL.
 DR HAMAP: MF_00149; -; 1
 DR InterPro: IPR003594; ATPbind_ATPase.
 DR InterPro: IPR002099; DNA_mis_repair.
 DR Pfam: PF01119; DNA_mis_repair; 1.
 DR Pfam: PF02518; HATPase_C; 1.
 DR SMART: SM00387; HATPase_C; 1.
 DR TIGRPFAM: TIGR00585; mutL; 1.
 DR PROSITE: PS00058; DNA_MISMATCH_REPAIR_1; 1.
 DR DNA repair; Complete proteome.
 KW PROSITE: PS00058; DNA_MISMATCH_REPAIR_1; 1.
 SQ SEQUENCE 618 AA; 67762 MW; 2896CA26AD358584 CRC64;

Query Match 2.5%; Score 8; DB 1; Length 618;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 223 ALASISSV 230
 Db 98 ALASISSV 105

RESULT 7

MUTL_CHLFE STANDARD; PRT; 624 AA.
 AC O8KAX3;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE DNA mismatch repair protein mutL.
 GN MUTL OR CT2028.
 OS Chlorobium tepidum.
 OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
 CC Chlorobium.
 CX NCBI_TaxID=1097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TLS / ATCC 49652 / DSM 12025;
 RX MEDLINE=22103685; PubMed=12093901;
 RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
 RA Dodson R.J., Deboy R., Gwin M.L., Nelson W.C., Haft D.H.,
 RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
 RA Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
 RA Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,
 RA Vanatvean J., Khouli H., White O., Gruber T.M., Ketchum K.A.,
 RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;
 RT "The complete genome sequence of Chlorobium tepidum TLS, a
 RT photosynthetic, anaerobic, green-sulfur bacterium";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).

CC -1- FUNCTION: This protein is involved in the repair of mismatches in
 CC DNA. It is required for dam-dependent methyl-directed DNA mismatch
 CC repair. May act as a "molecular matchmaker", a protein that
 CC promotes the formation of a stable complex between two or more
 CC DNA-binding proteins in an ATP-dependent manner without itself
 CC being part of a final effector complex (By similarity).
 CC -1- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; AE012953; AAW73245.1; -
 CC TIGR; CT2028; -
 DR HAMAP; MF_00149; -; 1.
 DR InterPro; IPR003594; ATPbind_ATPase.
 DR InterPro; IPR002099; DNA_mis_repair.
 DR Pfam; PF01119; DNA_mis_repair; 1.
 DR Pfam; PF02518; HATPase_C; 1.
 DR TIGRFAMs; TIGR00585; mutL; 1.
 DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
 KW DNA repair; Complete proteome.
 SQ SEQUENCE 624 AA; 69836 MW; 43F3CE8234A42CF6 CRC64;

Query Match 2.5%; Score 8; DB 1; Length 624;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 223 ALASISSV 230
 Db 99 ALASISSV 106

RESULT 8

MUTL_HABIN STANDARD; PRT; 629 AA.
 AC P44934;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE DNA mismatch repair protein mutL.
 GN MUTL OR HI0067.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 CC Pasteurellaceae; Haemophilus.
 CX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 RX MEDLINE=9530630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bilt C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudke D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fritchman U.L., Geoghegan N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RT Rd.";

CC -1- FUNCTION: This protein is involved in the repair of mismatches in
 CC DNA. It is required for dam-dependent methyl-directed DNA mismatch
 CC repair. May act as a "molecular matchmaker", a protein that
 CC promotes the formation of a stable complex between two or more
 CC DNA-binding proteins in an ATP-dependent manner without itself
 CC being part of a final effector complex (By similarity).
 CC -1- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; U32692; AAC21745.1; -
 DR PIR; B64046; E64046.
 DR HSRP; P23367; IBKN.
 DR TIGR; HI0067; -
 DR HAMAP; MF_00149; -; 1.
 DR InterPro; IPR003594; ATPbind_ATPase.
 DR InterPro; IPR002099; DNA_mis_repair.
 DR Pfam; PF01119; DNA_mis_repair; 1.
 DR Pfam; PF02518; HATPase_C; 1.
 DR SMART; SM00387; HATPase_C; 1.
 DR TIGRFAMs; TIGR00585; mutL; 1.
 DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
 KW DNA repair; Complete proteome.
 SQ SEQUENCE 629 AA; 71622 MW; 0AD240FD9AD1556 CRC64;

Query Match 2.5%; Score 8; DB 1; Length 629;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 223 ALASISSV 230
 Db 98 ALASISSV 105

RESULT 9

MUTL_PSEPK STANDARD; PRT; 632 AA.
 AC O88D11;
 ID MUTL_PSEPK

```

DT 15-MAR-2004 (Rel. 43, Last Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE DNA mismatch repair protein mutL.
GN MUTL OR PA4896.
OS Pseudomonas putida (strain KT2440).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=160488;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22423060; PubMed=12534463;
RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA Marston K.E., V.A.P., Fouts D.B., Gill S.R., Pop M., Holmes M.,
RA Brinac L., Benam M., DeBoy R.T., Daugherty S., Kolonay J.,
RA Madupu R., Nelson W., White O., Peterson J., Khouli H., Hance I.,
RA Chris Lee P., Holtzaple E., Scannan D., Tran K., Moazzez A.,
RA Uteback T., Rizzo M., Lee K., Kosack D., Mostl D., Medler H.,
RA Lauber J., Stjepandic D., Hohnselt J., Straetz M., Heim S.,
RA Kewitz C., Eisen J.A., Timmis K.N., Duesterhoeft A., Tsemmler B.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative analysis of the
RT metabolically versatile Pseudomonas putida KT2440."
RL Environ. Microbiol. 4:799-808(2002).
CC -1- FUNCTION: This protein is involved in the repair of mismatches in
CC DNA. It is required for dam-dependent methyl-directed DNA mismatch
CC repair. May act as a "molecular matchmaker", a protein that
CC promotes the formation of a stable complex between two or more
CC DNA-binding proteins in an ATP-dependent manner without itself
CC being part of a final effector complex (By similarity).
CC -1- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AEO16792; AAN70463.1; -
CC TIGR; PP4896; -
DR HAMAP; MF_00149; -; 1.
DR InterPro; IPR003594; ATPbind_ATPase.
DR InterPro; IPR002099; DNA_mis_repair.
DR Pfam; PF02518; HATPase_C; 1.
DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
KM DNA repair; Complete proteome.
SQ SEQUENCE 632 AA; 69745 MW; F7B93ECF5F7BAD8F CRC64;

Query Match 2.5%; Score 8; DB 1; Length 632;
Best Local Similarity 100.0%; Pred.No.12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 223 ALASISVS 230
DB 102 ALASISVS 109

RESULT 10
MUTL_PSEAE STANDARD; PRT; 633 AA.
ID MUTL_PSEAE
AC O9HUT8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA mismatch repair protein mutL.
GN MUTL OR PA4946.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warriner P.,
RA Hickey M.J., Brickman F.S.L., Hurnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.A., Coulter S.N., Folger K.R., Kae A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sajer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
CC -1- FUNCTION: This protein is involved in the repair of mismatches in
CC DNA. It is required for dam-dependent methyl-directed DNA mismatch
CC repair. May act as a "molecular matchmaker", a protein that
CC promotes the formation of a stable complex between two or more
CC DNA-binding proteins in an ATP-dependent manner without itself
CC being part of a final effector complex (By similarity).
CC -1- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AEO04907; AAC08331.1; -
CC PIR; P83028; F83028.
DR HAMAP; MF_00149; -; 1.
DR InterPro; IPR003594; ATPbind_ATPase.
DR InterPro; IPR002099; DNA_mis_repair.
DR Pfam; PF02518; HATPase_C; 1.
DR SMART; SM00387; HATPase_C; 1.
DR TIGRPFAM; TIGR00585; mutL; 1.
DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
KM DNA repair; Complete proteome.
SQ SEQUENCE 633 AA; 69747 MW; 3FD58AF0A721FE2D CRC64;

Query Match 2.5%; Score 8; DB 1; Length 633;
Best Local Similarity 100.0%; Pred.No.12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 223 ALASISVS 230
DB 102 ALASISVS 109

RESULT 11
MUTL_YERPE STANDARD; PRT; 635 AA.
ID MUTL_YERPE
AC O8ZIN4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA mismatch repair protein mutL.
GN MUTL OR YPO0371 OR Y0628.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Bivovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,
RA Penrice M.B., Sebald M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,

```

RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
 RA Lechner S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.,
 RA Simmonds M., Skelton U., Stevens K., Whitehead S., Barrett B.G.,
 RT "Genome sequence of Yersinia pestis, the causative agent of plague.",
 RL Nature 413:523-527(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KIMS / Biovar Mediaevalis;
 RX MEDLINE=22137863; PubMed=12142430;
 RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
 RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
 RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
 RA Straley S.C., McDonough K.A., Nilles M.L., Watson J.S., Blattner F.R.,
 RA Perry R.D.,
 RT "Genome sequence of Yersinia pestis KIM.",
 RL J. Bacteriol. 184:4601-4611(2002).
 CC -1- FUNCTION: This protein is involved in the repair of mismatches in
 CC DNA. It is required for dam-dependent methyl-directed DNA mismatch
 CC repair. May act as a "molecular matchmaker", a protein that
 CC promotes the formation of a stable complex between two or more
 CC DNA-binding proteins in an ATP-dependent manner without itself
 CC being part of a final effector complex (By similarity).
 CC -1- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AJ414142; CAC89230.1; -
 CC DR EMBL: AB013665; AAM84216.1; -
 CC DR PIR: AC0046; AC0046.
 CC DR HAMAP: MF_00149; -; 1.
 CC DR InterPro: IPR003594; ATPbind_ATPase.
 CC DR InterPro: IPR002099; DNA_mis_repair.
 CC DR Pfam: PF01119; DNA_mis_repair; 1.
 CC DR Pfam: PF02518; HATPase_C; 1.
 CC DR SMART: SMO0387; HATPase_C; 1.
 CC DR TIGRFAM: TIGR00585; mutL; 1.
 CC DR PROSITE: PS00058; DNA_MISMATCH_REPAIR_1; 1.
 CC KW DNA repair; Complete proteome.
 CC SEQUENCE 635 AA; 70287 MW; 301CAB1150449D45 CRC64;
 SQ
 Query Match 2.5%; Score 8; DB 1; Length 635;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 223 ALASISSV 230
 DB 98 ALASISSV 105
 RESULT 12
 MUTL_PSESMS
 ID MUTL_PSESMS STANDARD; PRT; 645 AA.
 AC 087072;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE DNA mismatch repair protein mutL.
 GN MUTL OR PSPRO4944.
 OS Pseudomonas syringae (pv. tomato).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=323;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DC3000;
 RX MEDLINE=22834015; PubMed=12928499;
 RA Buell C.R., Joerding V., Lindeberg M., Selengut J., Paulsen I.T.,

RA Gwin M.L., Dodson R.J., Depoy R.T., Durkin A.S., Kolonay J.F.,
 RA Madupu R., Daugherty S., Brinkac L., Beanan M.J., Haft D.H.,
 RA Nelson M.C., Davidson T., Zafar N., Zhou L., Liu J., Yuan Q.,
 RA Knorr H., Fedorova N., Tran B., Russell D., Berry K., Uetzel T.,
 RA Khan S.E., Feldblyum T.V., D'Ascenzo M., Deng W.-L., Ramos A.R.,
 RA Alfano J.R., Carlinhour S., Chatterjee A.K., Delaney T.P.,
 RA Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X., Bender C.L.,
 RA White O., Fraser C.M., Collier A.,
 RT "The complete genome sequence of the Arabidopsis and tomato pathogen
 RT Pseudomonas syringae pv. tomato DC3000.",
 RL Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).
 CC -1- FUNCTION: This protein is involved in the repair of mismatches in
 CC DNA. It is required for dam-dependent methyl-directed DNA mismatch
 CC repair. May act as a "molecular matchmaker", a protein that
 CC promotes the formation of a stable complex between two or more
 CC DNA-binding proteins in an ATP-dependent manner without itself
 CC being part of a final effector complex (By similarity).
 CC -1- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE016873; AA058372.1; -
 CC DR TIGR: PSPRO4944; -
 CC DR HAMAP: MF_00149; -; 1.
 CC DR InterPro: IPR003594; ATPbind_ATPase.
 CC DR InterPro: IPR002099; DNA_mis_repair.
 CC DR Pfam: PF01119; DNA_mis_repair; 1.
 CC DR Pfam: PF02518; HATPase_C; 1.
 CC DR PROSITE: PS00058; DNA_MISMATCH_REPAIR_1; 1.
 CC KW DNA repair; Complete proteome.
 CC SEQUENCE 645 AA; 70972 MW; 58245DE1B44515F CRC64;
 SQ
 Query Match 2.5%; Score 8; DB 1; Length 645;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 223 ALASISSV 230
 DB 118 ALASISSV 125
 RESULT 13
 MUTL_STAEP
 ID MUTL_STAEP STANDARD; PRT; 645 AA.
 AC 08CPB9;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE DNA mismatch repair protein mutL.
 GN MUTL OR SE0975.
 OS Staphylococcus epidermidis.
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=1282;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 12228;
 RX PubMed=12950922;
 RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
 RA Qin Z.-Q., Zhao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,
 RA Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.,
 RT "Genome-based analysis of virulence genes in a non-biofilm-forming
 RT Staphylococcus epidermidis strain (ATCC 12228).",
 RL Mol. Microbiol. 49:1577-1593(2003).
 CC -1- FUNCTION: This protein is involved in the repair of mismatches in
 CC DNA. It is required for dam-dependent methyl-directed DNA mismatch
 CC repair. May act as a "molecular matchmaker", a protein that
 CC promotes the formation of a stable complex between two or more

```

CC      DNA-binding proteins in an ATP-dependent manner without itself
CC      being part of a final effector complex (By similarity).
CC      -!- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; AE016747; AAC04572.1; -
CC      InterPro; IPR003594; ATPbind_ATPase.
CC      InterPro; IPR002099; DNA_mis_repair.
CC      Pfam; PF01119; DNA_mis_repair; 1.
CC      TIGRPFAMs; TIGR00585; mutL; 1.
CC      PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
CC      DNA repair; Complete proteome.
CC      SEQUENCE 645 AA; 75526 MW; FPA17BD6C3E47DC9 CRC64;

Query Match          2.5%; Score 8; DB 1; Length 645;
Best Local Similarity 100.0%; Pred.No.12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      223 ALA5ISSV 230
DB      99 ALA5ISSV 106

RESULT 14
MUTL_VIBCH STANDARD; PRT; 653 AA.
ID      MUTL_VIBCH STANDARD; PRT; 653 AA.
AC      Q9KAT3;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      DNA mismatch repair protein mutL.
OS      MUTL OR VC0345.
NC      Vibrrio cholerae.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC      Vibrionaceae; Vibrrio.
OX      NCBI_TaxID=666;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=El Tor N16961 / Serotype O1;
RX      MEDLINE=20406833; PubMed=10952301;
RA      Heidelberg U.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA      Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayen L.A.,
RA      Gill S.R., Nelson K.E., Read T.D., Teichlin H., Richardson D.,
RA      McAnuleva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA      McDonald L., Urdach T., Fleischmann R.D., Nierman W.C., White O.,
RA      Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA      Fraser C.M.;
RT      "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT      cholerae.";
RL      Nature 406:477-483(2000).
CC      -!- FUNCTION: This protein is involved in the repair of mismatches in
CC      DNA. It is required for dam-dependent methyl-directed DNA mismatch
CC      repair. May act as a "molecular matchmaker", a protein that
CC      promotes the formation of a stable complex between two or more
CC      DNA-binding proteins in an ATP-dependent manner without itself
CC      being part of a final effector complex (By similarity).
CC      -!- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).

```

```

CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; AE004123; AA993518.1; -
CC      FIR; A82354; A82354.
CC      HSP; P23567; IBKN.
CC      TIGR; VC0345; -
CC      HAMAP; MF 00149; -; 1.
CC      InterPro; IPR003594; ATPbind_ATPase.
CC      InterPro; IPR002099; DNA_mis_repair.
CC      Pfam; PF01119; DNA_mis_repair; 1.
CC      SMART; SM00387; ATPase_C; 1.
CC      TIGRPFAMs; TIGR00585; mutL; 1.
CC      PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
CC      DNA repair; Complete proteome.
CC      SEQUENCE 653 AA; 72630 MW; 5D970C21A17C66FF CRC64;

Query Match          2.5%; Score 8; DB 1; Length 653;
Best Local Similarity 100.0%; Pred.No.13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      223 ALA5ISSV 230
DB      98 ALA5ISSV 105

RESULT 15
MUTL_VIBVU STANDARD; PRT; 664 AA.
ID      MUTL_VIBVU STANDARD; PRT; 664 AA.
AC      Q8DCV0;
DT      10-OCT-2003 (Rel. 42, Created)
DT      10-OCT-2003 (Rel. 42, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      DNA mismatch repair protein mutL.
OS      MUTL OR VV11292.
NC      Vibrio vulnificus.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC      Vibrionaceae; Vibrrio.
OX      NCBI_TaxID=672;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CMCP6;
RA      Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA      Choy H.E.;
RT      "Complete genome sequence of Vibrio vulnificus CMCP6."
RT      Submitted (DEC-2002) to the EMBL/Genbank/DBJ databases.
CC      -!- FUNCTION: This protein is involved in the repair of mismatches in
CC      DNA. It is required for dam-dependent methyl-directed DNA mismatch
CC      repair. May act as a "molecular matchmaker", a protein that
CC      promotes the formation of a stable complex between two or more
CC      DNA-binding proteins in an ATP-dependent manner without itself
CC      being part of a final effector complex (By similarity).
CC      -!- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; AE016801; AAC09747.1; -
CC      HAMAP; MF 00149; -; 1.
CC      InterPro; IPR003594; ATPbind_ATPase.
CC      InterPro; IPR002099; DNA_mis_repair.
CC      Pfam; PF01119; DNA_mis_repair; 1.
CC      SMART; SM00387; ATPase_C; 1.
CC      TIGRPFAMs; TIGR00585; mutL; 1.
CC      PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
CC      DNA repair; Complete proteome.
CC      SEQUENCE 664 AA; 74006 MW; A0ECF5F88247A7C CRC64;

```

Query Match 2.5%; Score 8; DB 1; Length 664;
Best Local Similarity 100.0%; Pred.No.13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
|||
Db 98 ALASISSV 105

Search completed: February 23, 2004, 19:33:49
Job time : 19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2004, 19:19:21 ; Search time 59 seconds

(without alignments)
1532.461 Million cell updates/sec

Title: US-09-936-456-2

Perfect score: 320

Sequence: 1 MTLWNGVLPFPYPPRRHAGF.....GDPHKQALPDKCTITVL 320

Scoring table: OLIGO

Searched: 1586107 seqs, 282547505 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: A_Geneseq_29Jan04:*

1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	320	100.0	320	3	AA18992 Amino aci
2	320	100.0	320	3	AA26325 Human CAS
3	220	68.8	242	3	AA52095 Human sec
4	219	68.4	320	5	AB99126 Human pol
5	219	68.4	370	7	ADD19303 Human sec
6	172	53.8	262	3	AA26326 CASB618 p
7	49	15.3	103	7	ADD19262 Human sec
8	19	5.9	98	4	AA26378 Human col
9	19	5.9	112	4	AA26373 Human INT
10	19	5.9	112	4	ABO32727 Secreted
11	19	5.9	298	6	AD25266 Human pro
12	19	5.9	343	4	AA26371 Human INT
13	19	5.9	343	6	ABU08370 Amino aci
14	19	5.9	445	6	ABO32725 Secreted
15	19	5.9	445	4	ABO32725 Novel hum
16	13	4.1	343	5	AB26378 Human CAS
17	10	3.1	10	3	AA26397 Human CAS
18	10	3.1	10	3	AA26399 Human CAS
19	10	3.1	10	3	AA26371 Human CAS
20	10	3.1	10	3	AA26381 Human CAS
21	10	3.1	10	3	AA26396 Human CAS
22	10	3.1	10	3	AA26369 Human CAS
23	10	3.1	10	3	AA26373 Human CAS
24	10	3.1	10	3	AA26382 Human CAS
25	10	3.1	10	3	AA26382 Human CAS

26	10	3.1	10	3	AA26379 Human CAS
27	10	3.1	10	3	AA26392 Human CAS
28	10	3.1	10	3	AA26375 Human CAS
29	10	3.1	10	3	AA26384 Human CAS
30	10	3.1	10	3	AA26389 Human CAS
31	10	3.1	10	3	AA26370 Human CAS
32	10	3.1	10	3	AA26372 Human CAS
33	10	3.1	10	3	AA26376 Human CAS
34	10	3.1	10	3	AA26387 Human CAS
35	10	3.1	10	3	AA26385 Human CAS
36	10	3.1	10	3	AA26385 Human CAS
37	10	3.1	10	3	AA26385 Human CAS
38	10	3.1	10	3	AA26374 Human CAS
39	10	3.1	10	3	AA26393 Human CAS
40	10	3.1	10	3	AA26386 Human CAS
41	10	3.1	10	3	AA26398 Human CAS
42	10	3.1	10	3	AA26377 Human CAS
43	10	3.1	10	3	AA26380 Human CAS
44	10	3.1	10	3	AA26390 Human CAS
45	10	3.1	10	3	AA26391 Human CAS

ALIGNMENTS

RESULT 1	AA18992 standard; protein; 320 AA.
ID	AA18992
XX	AA18992:
AC	08-FEB-2001 (first entry)
DT	
XX	
DE	Amino acid sequence of a human transmembrane protein.
XX	
KW	Human; transmembrane protein; cell proliferation disorder; myeloma;
KW	reproductive disorder; smooth muscle disorder; neurological disorder;
KW	arteriosclerosis; leukemia; acquired immunodeficiency syndrome; AIDS;
KW	allergy; ovulatory defect; angina; hypertension; stroke; epilepsy;
KW	Alzheimer's disease; Tourette's disorder.
XX	
OS	Homo sapiens.
XX	
PH	
FT	Key
FT	Location/Qualifiers
FT	Modified-site
FT	84 /note= "potential glycosylation site"
FT	86 /note= "potential phosphorylation site"
FT	96 /note= "potential phosphorylation site"
FT	109 /note= "potential phosphorylation site"
FT	121 /note= "potential glycosylation site"
FT	131 /note= "potential glycosylation site"
FT	156 /note= "potential phosphorylation site"
FT	156 /note= "potential glycosylation site"
FT	280 /note= "potential phosphorylation site"
FT	285 /note= "potential phosphorylation site"
FT	Modified-site
FT	285 /note= "potential phosphorylation site"
XX	
PD	WO20056891-A2.
XX	
PF	28-SEP-2000.
XX	
PR	22-MAR-2000; 2000WO-US007817.
XX	
PR	22-MAR-1999; 99US-0125537P.
XX	
PR	16-JUN-1999; 99US-0139565P.
XX	
PA	(INCY-) INCYTE PHARM INC.

XX Yue H, Lai P, Tang YT, Hillman JL, Reddy R, Bandman O;
 PI Baughn MR, Lu DM, Azimzai Y, Yang J;
 XX WPI; 2000-579485/54.
 DR N-PSDB; AAA96505.
 XX
 PT New human transmembrane proteins are used to treat a disease or condition
 associated with decreased expression of functional HMP e.g. Tourette's
 disorder, angina and leukemia.
 XX
 PS Claim 1; Page 109, 130pp; English.
 XX
 CC The present sequence represents a human transmembrane proteins (HMP).
 CC Agonists and antagonists of the protein are used to treat a disease or
 CC condition associated with overexpression of the protein. Diseases and
 CC conditions which can be treated include cell proliferative,
 CC immunological, reproductive, smooth muscle and neurological disorders
 CC e.g. arteriosclerosis, myeloma, leukaemia, acquired immunodeficiency
 CC syndrome (AIDS), allergies, ovulatory defects, angina, hypertension,
 CC stroke, Alzheimer's disease, epilepsy and Tourette's disorder. The
 CC polynucleotides may be used to detect and quantify gene expression in
 CC biopsied tissues where protein expression may be correlated with disease
 CC e.g. to determine absence, presence or excess expression of HMP or to
 CC monitor regulation of HMP expression during therapeutic intervention
 XX
 SQ Sequence 320 AA;
 Query Match 100.0%; Score 320; DB 3; Length 320;
 Best Local Similarity 100.0%; Pred. No. 1.2e-300;
 Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTLMNGVLPFYPPQPHAAAGFSVPLILVILVFLAALASFLILPGRGSRMFWLVYLLS 60
 DB 1 MTLMNGVLPFYPPQPHAAAGFSVPLILVILVFLAALASFLILPGRGSRMFWLVYLLS 60
 QY 61 LFIGAEIVAVHPSAEFVGTNTNTSYKAFSAARVTARVGLVGLGEGINITLTGTPVHQL 120
 DB 61 LFIGAEIVAVHPSAEFVGTNTNTSYKAFSAARVTARVGLVGLGEGINITLTGTPVHQL 120
 QY 121 NETIDYNEQFTWRLKENYAAEYANALEKGLPDPVLYLAKEFTPSPCGLYHQYHLAghYA 180
 DB 121 NETIDYNEQFTWRLKENYAAEYANALEKGLPDPVLYLAKEFTPSPCGLYHQYHLAghYA 180
 QY 181 SATLWVAFCFWILSNVLSTPAPLYGGLALLTTGAFALFGVPALASISVPLCPRLGSS 240
 DB 181 SATLWVAFCFWILSNVLSTPAPLYGGLALLTTGAFALFGVPALASISVPLCPRLGSS 240
 QY 241 ALTTQYGAAFWVTLATGVLCEFLGAVVSLQYVPSALRTLLDQSAKDCQSERGSPILL 300
 DB 241 ALTTQYGAAFWVTLATGVLCEFLGAVVSLQYVPSALRTLLDQSAKDCQSERGSPILL 300
 QY 301 GDPHLKQALPDLKICITTNL 320
 DB 301 GDPHLKQALPDLKICITTNL 320
 RESULT 2
 ID AAB26325
 AAB26325 standard; protein; 320 AA.
 AC AAB26325;
 DT 11-JAN-2001 (first entry)
 XX Human CASB618 protein.
 DE Human CASB618 protein.
 KW Epitope; human; CASB618; chromosome 15; tumour; vaccine; cancer; ovarian;
 XX colon; autoimmune disease; HLA-A0201.
 OS Homo sapiens.
 XX
 PN W0200053748-A2.

XX 14-SEP-2000.
 PD
 XX 09-MAR-2000; 2000WO-EP002048.
 PF
 XX 11-MAR-1999; 99GB-00005607.
 PR 01-SEP-1999; 99GB-00020590.
 XX
 XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 PA
 XX Bruck CEM, Cassart J, Coche T, Vinals Y De BassolsJC;
 PI
 XX WPI; 2000-572268/53.
 DR N-PSDB; AAA94623.
 XX
 PT New human CASB618 polypeptide, useful as a vaccine for prophylactic and
 PT therapeutic treatment of cancers, particularly ovarian or colon cancer,
 PT autoimmune diseases and related conditions.
 XX
 PS Claim 1; Page 61-62; 76pp; English.
 XX
 CC The present sequence is human CASB618 protein. The gene for human CASB618
 CC is thought to be located on chromosome 15. The present protein and
 CC epitopes of this protein (see AAB26327 to AAB26399) are useful in
 CC diagnosing the occurrence of tumour cells and in vaccines for prophylactic
 CC and therapeutic treatment of cancers, particularly ovarian or colon
 CC cancer, autoimmune diseases and related conditions
 XX
 SQ Sequence 320 AA;
 Query Match 100.0%; Score 320; DB 3; Length 320;
 Best Local Similarity 100.0%; Pred. No. 1.2e-300;
 Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTLMNGVLPFYPPQPHAAAGFSVPLILVILVFLAALASFLILPGRGSRMFWLVYLLS 60
 DB 1 MTLMNGVLPFYPPQPHAAAGFSVPLILVILVFLAALASFLILPGRGSRMFWLVYLLS 60
 QY 61 LFIGAEIVAVHPSAEFVGTNTNTSYKAFSAARVTARVGLVGLGEGINITLTGTPVHQL 120
 DB 61 LFIGAEIVAVHPSAEFVGTNTNTSYKAFSAARVTARVGLVGLGEGINITLTGTPVHQL 120
 QY 121 NETIDYNEQFTWRLKENYAAEYANALEKGLPDPVLYLAKEFTPSPCGLYHQYHLAghYA 180
 DB 121 NETIDYNEQFTWRLKENYAAEYANALEKGLPDPVLYLAKEFTPSPCGLYHQYHLAghYA 180
 QY 181 SATLWVAFCFWILSNVLSTPAPLYGGLALLTTGAFALFGVPALASISVPLCPRLGSS 240
 DB 181 SATLWVAFCFWILSNVLSTPAPLYGGLALLTTGAFALFGVPALASISVPLCPRLGSS 240
 QY 241 ALTTQYGAAFWVTLATGVLCEFLGAVVSLQYVPSALRTLLDQSAKDCQSERGSPILL 300
 DB 241 ALTTQYGAAFWVTLATGVLCEFLGAVVSLQYVPSALRTLLDQSAKDCQSERGSPILL 300
 QY 301 GDPHLKQALPDLKICITTNL 320
 DB 301 GDPHLKQALPDLKICITTNL 320
 RESULT 3
 ID AAB52095
 AAB52095 standard; protein; 242 AA.
 AC AAB52095;
 DT 21-FEB-2001 (first entry)
 XX Human secreted protein sequence encoded by gene 44 SEQ ID NO:144.
 DE Human secreted protein sequence encoded by gene 44 SEQ ID NO:144.
 KW Human; secreted protein; cytostatic; immunosuppressive; nootropic;
 KW neuroprotective; antiviral; antiallergic; hepatocytic; antidiabetic;
 KW antiinflammatory; antiviral; vulnerable; anticonvulsant; antibacterial;
 KW antifungal; antiparasitic; cardiac; cancer; immune disease; allergy;

QY 121 NETIDYNEQFTWRLKENYAAYANALEKGLPDPVLYLAKEFTPSSPCGLHYQYHLAghya 180
 CC |
 CC |
 Db 121 NETIDYNEQFTWRLKENYAAYANALEKGLPDPVLYLAKEFTPSSPCGLHYQYHLAghya 180
 CC |
 CC |
 QY 181 SATLWVAFPCFWLLSNVLLSTPAFLYVGGIALTTGAPALFGVPAALASISSVPLCPRLGSS 240
 CC |
 CC |
 Db 181 SATLWVAFPCFWLLSNVLLSTPAFLYVGGIALTTGAPALFGVPAALASISSVPLCPRLGSS 240
 CC |
 CC |
 QY 241 ALTTQYGAAFWVTLATGVCLFLGGAVVSLOYRPSALRTLLDQSAKCCSGERGSSPLL 300
 CC |
 CC |
 Db 241 ALTTQYGAAFWVTLATGVCLFLGGAVVSLOYRPSALRTLLDQSAKCCSGERGSSPLL 300
 CC |
 CC |
 QY 301 GDP LHKQALPDLKCIITNNL 320
 CC |
 CC |
 Db 301 GDP LHKQALPDLKCIITNNL 320
 CC |
 CC |
 RESULT 5
 ADD19303
 ID ADD19303 standard; protein; 370 AA.
 XX
 AC ADD19303;
 XX
 DT 15-JAN-2004 (first entry)
 XX
 DE Human secreted protein from gene 4 #2.
 XX
 KW human secreted protein; cytostatic; antibacterial; virucide;
 KW neuroprotective; gynaecological; gastroenterological; cardiac;
 KW cardiovascular; gen; nephrotoxic; antiinflammatory; muscular; gen;
 KW respiratory; gen; immunosuppressive; cerebroprotective; vasotropic;
 KW neutropilic; antiallergic; cancer; bacterial infection; viral infection;
 KW neutral disorder; immune system disorder; blood disorder;
 KW muscular disorder; reproductive disorder; gastrointestinal disorder;
 KW pulmonary disorder; cardiovascular disorder; renal disorder;
 KW inflammatory disorder; proliferative disorder; human.
 XX
 OS Homo sapiens.
 XX
 PN W02003052377-A2.
 XX
 PD 26-JUN-2003.
 XX
 PF 06-NOV-2002; 2002WC-US036606.
 XX
 PR 07-NOV-2001; 2001US-0331046P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM,
 XX
 DR WPI: 2003-533050/50.
 DR N-PSDB; ADD19228.
 XX
 PT New isolated nucleic acids encoding signal transduction pathway component
 PT polypeptides, useful for diagnosing, treating, and/or preventing
 PT disorders, such as cancer, infections, cardiovascular and inflammatory
 PT diseases.
 XX
 PS Claim 11; SEQ ID NO 130; 554pp; English.
 XX
 CC The invention relates to an isolated nucleic acid molecule (cDNA)
 CC encoding a human secreted protein, representing one of 85 novel genes.
 CC Also included are recombinant vectors, host cells (expressing the
 CC protein), the secreted proteins (including their fragments, epitopes and
 CC homologues), an isolated antibody that binds specifically to the protein,
 CC diagnosing a pathological condition or susceptibility to a pathological
 CC condition (comprising determining the presence or absence of a mutation
 CC in the nucleic acid and diagnosing a pathological condition or
 CC absence of the mutation), diagnosing a pathological condition or
 CC susceptibility to a pathological condition (comprising determining the
 CC presence or amount of expression of the protein in a biological sample
 CC and diagnosing a condition based on the presence or amount of expression

CC of the protein), preventing, treating or ameliorating a medical condition
 CC by administering the nucleic acid or protein to a mammalian subject,
 CC identifying a binding partner to the protein, the gene corresponding to
 CC the cDNA sequence, and identifying an activity in a cell, isolating the
 CC (comprising expressing the nucleic acid in a cell, isolating the
 CC supernatant, detecting an activity in a biological assay and identifying
 CC the protein in the supernatant having the activity). The nucleic acids
 CC and proteins display the following activities: cytostatic, antibacterial,
 CC virucide, neuroprotective, gynaecological, gastroenterological, gen,
 CC cardiac, cardiovascular, gen, nephrotoxic, antiinflammatory, muscular,
 CC neutropilic, antiallergic. The methods and compositions of the present
 CC invention are useful for diagnosing, treating, preventing and/or
 CC prognosticating disorders related to the novel polypeptides, such as
 CC cancer, bacterial or viral infections, and neutral, immune system, blood,
 CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
 CC renal, inflammatory or proliferative disorders (many examples of these
 CC diseases and disorders are given in the specification). The present
 CC sequence represents a novel secreted protein of the invention.
 CC
 XX
 SQ Sequence 370 AA;
 XX
 Query Match 68.4%; Score 219; DB 7; Length 370;
 Best Local Similarity 99.7%; Pred. No. 6,9e-203;
 Matches 319; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MTLWNGVLPFPQPHAAFGSVPLIVLVLALAAFPLLIPGIRGSHFWMLVRYLLS 60
 Db 51 MTLWNGVLPFPQPHAAFGSVPLIVLVLALAAFPLLIPGIRGSHFWMLVRYLLS 110
 QY 61 LFTGAELVAVHPSAEMFGTNTNTSYAFSAFARVTLRVGLVGLGINTLLTGPVHOL 120
 Db 111 LFTGAELVAVHPSAEMFGTNTNTSYAFSAFARVTLRVGLVGLGINTLLTGPVHOL 170
 QY 121 NETIDYNEQFTWRLKENYAAYANALEKGLPDPVLYLAKEFTPSSPCGLHYQYHLAghya 180
 Db 171 NETIDYNEQFTWRLKENYAAYANALEKGLPDPVLYLAKEFTPSSPCGLHYQYHLAghya 230
 QY 181 SATLWVAFPCFWLLSNVLLSTPAFLYVGGIALTTGAPALFGVPAALASISSVPLCPRLGSS 240
 Db 231 SATLWVAFPCFWLLSNVLLSTPAFLYVGGIALTTGAPALFGVPAALASISSVPLCPRLGSS 290
 QY 241 ALTTQYGAAFWVTLATGVCLFLGGAVVSLOYRPSALRTLLDQSAKCCSGERGSSPLL 300
 Db 291 ALTTQYGAAFWVTLATGVCLFLGGAVVSLOYRPSALRTLLDQSAKCCSGERGSSPLL 350
 QY 301 GDP LHKQALPDLKCIITNNL 320
 Db 351 GDP LHKQALPDLKCIITNNL 370
 RESULT 6
 AAB26326
 ID AAB26326 standard; protein; 262 AA.
 XX
 AC AAB26326;
 XX
 DT 12-SEP-2003 (revised)
 DT 11-JAN-2001 (first entry)
 XX
 DE CASB618 protein.
 XX
 KW Epitope; human; CASB618; chromosome 15; tumour; vaccine; cancer; ovarian;
 KW colon; autoimmune disease; HLA_A0201.
 XX
 OS Homo sapiens.
 OS Chimeric.
 OS
 PN W0200053748-A2.
 PD 14-SEP-2000.
 XX

PF 09-MAR-2000; 2000MO-EP002048.
XX 11-MAR-1999; 99GB-00005607.
PR 01-SEP-1999; 99GB-00020590.
XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX PA
XX PI Bruck CEM, Casasart J, Coche T, Vanals Y De Baesjoc;
XX WPI; 2000-572268/53.
DR
XX New human CASB618 polypeptide, useful as a vaccine for prophylactic and
XX therapeutic treatment of cancers, particularly ovarian or colon cancer,
XX autoimmune diseases and related conditions.
XX
XX Example 7; Page 63; 76pp; English.
XX
XX The present invention relates to epitopes of human CASB618 protein (see
XX AAB2637 to AAB2639). The epitopes of AAB2637 to AAB2639 are useful in
XX diagnosing the occurrence of tumour cells and in vaccines for prophylactic
XX and therapeutic treatment of cancers, particularly ovarian or colon
XX cancer, autoimmune diseases and related conditions. The present sequence
XX is a protein derived from a chimeric gene constructed from the CASB618
XX gene carrying deletions of the N-terminus and C-terminus, with the
XX addition of an ITP (NSI DNA sequence encoding the N-terminal 1-81 amino
XX acids of the NSI protein of Influenza virus) at the N-terminus, and a C-
XX terminal histidine tail. The gene encoding the present protein was cloned
XX and used to evaluate the effectiveness of tumour-specific antigens.
XX (Updated on 12-SEP-2003 to standardise OS field)
SQ Sequence 262 AA;

Query Match 53.8%; Score 172; DB 3; Length 262;
Best Local Similarity 100.0%; Pred. No. 1.4e-157;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 EMPVGTAVNTSYAFSAARVATAGVGLVGEINILTPPHQNETIDVNEOTWTL 134
DB 82 EMPVGTAVNTSYAFSAARVATAGVGLVGEINILTPPHQNETIDVNEOTWTL 141
QY 135 KENVAEYANALKEKLPDPVLYLAEKFTPSQCLYHQYHLAGHYASATLWAFCEWLLS 194
DB 142 KENVAEYANALKEKLPDPVLYLAEKFTPSQCLYHQYHLAGHYASATLWAFCEWLLS 201
QY 195 NVLSTPAPLYGGALLTTGAFALFGVAFALASISSVPLCPRLGSSALLTQY 246
DB 202 NVLSTPAPLYGGALLTTGAFALFGVAFALASISSVPLCPRLGSSALLTQY 253

RESULT 7
ADD19262
ID ADD19262 standard; protein; 103 AA.
XX
XX ADD19262;
XX
XX 15-JAN-2004 (first entry)
XX
XX Human secreted protein from gene 4.
DE
XX human secreted protein; cytosolic; antibacterial; virucide;
XX neutroprotective; gynaecological; gastrointestinal; Gen; cardiac;
XX cardiovascular; Gen; nephrotoxic; anti-inflammatory; muscular; Gen;
XX respiratory; Gen; immunosuppressive; cerebroprotective; vasotropic;
XX neurotoxic; anti-allergic; cancer; bacterial infection; viral infection;
XX neural disorder; immune system disorder; blood disorder;
XX muscular disorder; reproductive disorder; gastrointestinal disorder;
XX pulmonary disorder; cardiovascular disorder; renal disorder;
XX inflammatory disorder; proliferative disorder; human.
XX
XX Homo sapiens.
XX OS
XX PN WO2003052377-A2.
XX

PD 26-JUN-2003.
XX
XX 06-NOV-2002; 2002MO-US035606.
PF
XX 07-NOV-2001; 2001US-0331046P.
PR
XX (HUMA-) HUMAN GENOME SCI INC.
XX PA
XX PI Rosen CA, Ruben SM;
XX WPI; 2003-533050/50.
DR
XX N-PSDB; ADD19187.
XX
XX New isolated nucleic acids encoding signal transduction pathway component
XX polypeptides, useful for diagnosing, treating, and/or preventing
XX disorders, such as cancer, infections, cardiovascular and inflammatory
XX diseases.
XX
XX Claim 11; SEQ ID NO 89; 554bp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (cDNA)
XX encoding a human secreted protein, representing one of 85 novel genes.
XX Also included are recombinant vectors, host cells (expressing the
XX protein), the secreted proteins (including their fragments, epitopes and
XX homologues), an isolated antibody that binds specifically to the protein,
XX diagnosing a pathological condition or susceptibility to a pathological
XX condition (comprising determining the presence or absence of a mutation
XX in the nucleic acid and diagnosing a condition based on the presence or
XX absence of the mutation), diagnosing a pathological condition or
XX susceptibility to a pathological condition (comprising determining the
XX presence or amount of expression of the protein in a biological sample
XX and diagnosing a condition based on the presence or amount of expression
XX of the protein), preventing, treating or ameliorating a medical condition
XX by administering the nucleic acid or protein to a mammalian subject,
XX identifying a binding partner to the protein, the gene corresponding to
XX the cDNA sequence, and identifying an activity in a biological assay
XX (comprising expressing the nucleic acid in a cell, isolating the
XX supernatant, detecting an activity in a biological assay and identifying
XX the protein in the supernatant having the activity). The nucleic acids
XX and proteins display the following activities: cytotoxic, antibacterial,
XX virucide, neuroprotective, gynaecological, gastrointestinal, Gen,
XX cardiac, cardiovascular; Gen, nephrotoxic, anti-inflammatory, muscular;
XX Gen, respiratory; Gen, immunosuppressive, cerebroprotective, vasotropic,
XX neurotoxic, anti-allergic. The methods and compositions of the present
XX invention are useful for diagnosing, treating, preventing and/or
XX prognosticating disorders related to the novel polypeptides, such as
XX cancer, bacterial or viral infections, and neural, immune system, blood,
XX muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
XX renal, inflammatory or proliferative disorders (many examples of these
XX diseases and disorders are given in the specification). The present
XX sequence represents a novel secreted protein of the invention.
SQ Sequence 103 AA;

Query Match 15.3%; Score 49; DB 7; Length 103;
Best Local Similarity 100.0%; Pred. No. 5.3e-39;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTLNNGVLPFYPPORRAAGSVPLTIVTFLAASFLILPGIRGHS 49
DB 1 MTLNNGVLPFYPPORRAAGSVPLTIVTFLAASFLILPGIRGHS 49

RESULT 8
AAG76780
ID AAG76780 standard; protein; 98 AA.
XX
XX AAG76780;
XX
XX 03-SEP-2001 (first entry)
XX
XX Human colon cancer antigen protein SEQ ID NO:7544.
XX
XX

KM Human: colon cancer; colon cancer antigen; diagnosis; detection;
KM colorectal carcinoma.
XX
XX Homo sapiens.
OS
XX WO200122920-A2.
PN
XX
XX 05-APR-2001.
PD
XX 28-SEP-2000; 2000WO-US026524.
PF
XX 29-SEP-1999; 99US-0157137P.
PR 03-NOV-1999; 99US-0163280P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI WPI; 2001-235357/24.
XX
XX N-PSDB; AAH36185.
DR
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers.
XX
XX Claim 11; Page 8959-8960; 9803pp; English.
PS
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene therapy
CC and vaccine production. N and P may be used in the prevention, diagnosis
CC and treatment of diseases associated with inappropriate P expression. For
CC example, N and P may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of P by expressing inactive proteins or to
CC supplement the patient's own production of P. Additionally, N may be used
CC to produce the colon cancer-associated Pe, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the proteins. N and P
CC can be used in the prevention, diagnosis and treatment of colorectal
CC carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent
CC sequences used in the exemplification of the present invention. N.B.
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC time of publication, meaning no sequences are present for SEQ ID NO:1027
CC to 1052, 7921 and 7922
XX
XX Sequence 98 AA;
SQ
Query Match 5.9%; Score 19; DB 4; Length 98;
Best Local Similarity 100.0%; Pred. No. 5.2e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 50 RWFMLVRLVLSLFTGAEIV 68
DB 8 RWFMLVRLVLSLFTGAEIV 26
RESULT 9
AAB6073
ID AAB6073 standard; protein; 112 AA.
XX
XX AAB6073;
AC
XX 30-MAR-2001 (first entry)
DT
XX Human INTERCEPT 296 extracellular domain #2.
DE
XX TANGO protein; INTERCEPT protein; neurological disorder;
KW central nervous system; focal brain disorder; bipolar affective disorder;
KW global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;
KW senile dementia; Huntington's disease; amyotrophic lateral sclerosis;
KW Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;
KW neuropsychiatric; psychoactive substance use; anxiety.
XX

OS Homo sapiens.
XX
XX WO200077239-A2.
PN
XX
XX 21-DEC-2000.
PD
XX 24-MAY-2000; 2000WO-US014858.
PF
XX 14-JUN-1999; 99US-00333159.
PR
XX (MILL-) MILLENNIUM PHARM INC.
PA
XX McCarthy SA, Fraser CC, Sharp JD, Barnes TW;
PI WPI; 2001-032313/04.
XX
XX TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for
PT screening assays and diagnostic assays and for the treatment of
PT neurological diseases such as Alzheimer's, Parkinson's and Huntington's
PT disease.
XX
XX Claim 8; Page 331; 359pp; English.
PS
XX The present invention relates to TANGO or INTERCEPT proteins and coding
CC sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66031-B66057,
CC AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding
CC sequences are useful for the treatment of neurological disorders such as
CC central nervous system (CNS) disorders, CNS-related disorders, focal
CC brain disorders, global-diffuse cerebral disorders and other neurological
CC and cerebrovascular disorders. The CNS disorders include Alzheimer's
CC disease, senile dementia, Huntington's disease, amyotrophic lateral
CC sclerosis, Parkinson's, Gilles de la Tourette's syndrome, autonomic
CC function disorders such as hypertension and sleep disorders,
CC neuropsychiatric disorders, psychoactive substance use disorders,
CC anxiety, and bipolar affective disorder
XX
XX Sequence 112 AA;
SQ
Query Match 5.9%; Score 19; DB 4; Length 112;
Best Local Similarity 100.0%; Pred. No. 6e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 145 ALEKGLPDPVLYIAEKFTP 163
DB 75 ALEKGLPDPVLYIAEKFTP 93
RESULT 10
ABO32727
ID ABO32727 standard; protein; 112 AA.
XX
XX ABO32727;
AC
XX 17-SEP-2003 (first entry)
DT
XX Secreted polypeptide-related protein #101.
DE
XX Human: TANGO; INTERCEPT; secreted polypeptide; immune disorder;
KW hormonal disorder; proliferative disorder; cancer; thyroid disorder;
KW diabetes; multiple sclerosis; lupus; neurological disorder; anemia;
KW Alzheimer's disease; Parkinson's disease; cardiovascular disorder;
KW myocardial infarction; congestive heart disease; blood platelet disorder;
KW thrombocytopenia; blood vessel; atherosclerosis; vasculitis.
XX
XX Homo sapiens.
OS
XX US2003022279-A1.
PN
XX 30-JAN-2003.
PD
XX 12-JAN-2001; 2001US-00759130.
PF
XX 14-JUN-1999; 99US-00333159.
PR

PR 29-JUN-1999; 99US-00342364.
 FR 10-SEP-1999; 99US-00393996.
 FR 19-OCT-1999; 99US-00420707.
 PR 07-JAN-2000; 2000US-00479249.
 PR 27-APR-2000; 2000US-00559497.
 PR 24-MAY-2000; 2000US-00578063.
 PR 16-JUN-2000; 2000US-00596194.
 PR 23-JUN-2000; 2000US-00602871.
 PR 30-JUN-2000; 2000US-00608452.
 XX
 PA (FRAS/) FRASER C C.
 PA (BARN/) BARNES T M.
 PA (SHAR/) SHARP J D.
 PA (KIRS/) KIRST S J.
 PA (MYER/) MYERS P S.
 PA (LEIB/) LEIBY K R.
 PA (HOLT/) HOLTZMAN D A.
 PA (MCCA/) MCCARTHY S A.
 PA (WRIGHT/) WRIGHTON N.
 PA (MACK/) MACKAY C R.
 PA (GOOD/) GOODEARL A D J.
 PI Fraser CC, Barnes TM, Sharp JD, Kirst SJ, Myers PS, Leiby KR;
 PI Holtzman DA, McCarthy SA, Wrighton N, Mackay CR, Goodearl ADJ;
 XX WPI; 2003-456290/43.
 DR
 XX
 PT New nucleic acid molecule encoding a secreted protein (e.g. TANGO 202,
 PT TANGO 210 or INTERCEPT 217), useful for diagnosing, preventing or
 PT treating disorders such as cancer, diabetes or atherosclerosis, and in
 PT forensic biology.
 PT
 XX
 PS Claim 9; SEQ ID NO 427; 482bp; English.
 XX
 CC The invention relates to secreted polypeptide-related proteins and
 CC nucleic acids (TANGO and INTERCEPT proteins and nucleic acids). The
 CC nucleic acids, proteins and antibodies specific to the proteins are
 CC useful in screening assays, predictive medicine (e.g. diagnostic assays,
 CC prognostic assays, monitoring clinical trials and pharmacogenetics) and
 CC prophylactic and therapeutic methods. The sequences are used in
 CC diagnosing, preventing or treating proliferative disorders (e.g.
 CC cancers), hormonal disorders (e.g. diabetes or thyroid disorders), immune
 CC disorders (e.g. multiple sclerosis or lupus), neurological disorders
 CC (e.g. Alzheimer's disease or Parkinson's disease), cardiovascular
 CC disorders (e.g. myocardial infarction or congestive heart disease), blood
 CC platelet disorders (e.g. thrombocytopenia or anaemia) and disorders
 CC involving blood vessels (e.g. atherosclerosis or vasculitis). The nucleic
 CC acids may also be used in chromosome mapping, tissue typing and forensic
 CC biology, and as surrogate markers. This sequence represents a secreted
 CC polypeptide-related protein of the invention. Note: The sequence data for
 CC this patent was obtained in electronic format directly from USPTO at
 CC Seqdata.uspto.gov/sequence.html
 CC
 XX
 SQ Sequence 112 AA;
 Query Match 5.9%; Score 19; DB 6; Length 112;
 Best Local Similarity 100.0%; Pred. No. 6e-10; Mismatches 0; Indels 0; Gaps 0;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 145 ALEKGLPDPVLYAEKFTP 163
 Db 75 ALEKGLPDPVLYAEKFTP 93
 RESULT 11
 ID ADA55266 standard; protein; 298 AA.
 XX
 AC ADA55266;
 XX
 XX 20-NOV-2003 (first entry)
 DT
 XX Human protein, SEQ ID 2834.
 DE

XX
 KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
 KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
 KW Inflammatory disease; osteoporosis; neurological disease.
 XX
 OS Homo sapiens.
 XX
 XX EPI293565-A2.
 XX
 XX
 XX 19-MAR-2003.
 XX
 XX
 XX 21-MAR-2002; 2002BP-00006586.
 XX
 XX 14-SEP-2001; 2001JP-00328381.
 PR 24-JAN-2002; 2002US-0350435P.
 XX
 XX (HELI-) HELIX RES INST.
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.
 PA
 PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;
 XX WPI; 2003-395539/38.
 DR
 DR N-PSDB; ADA53627.
 XX
 PT New polynucleotides encoding full-length polypeptides, e.g. secretory
 PT and/or membrane proteins, useful for developing medicines for diseases in
 PT which the gene is involved, or as target molecules for gene therapy.
 PT
 XX
 PS Claim 14; SEQ ID NO 2834; 205bp; English.
 XX
 CC The present invention relates to novel human secretory or membrane
 CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
 CC ADA54071). The coding sequences are useful in the gene therapy of
 CC diseases caused by abnormalities of the proteins, e.g. cancer,
 CC inflammatory diseases, osteoporosis or neurological disease.
 CC
 XX
 SQ Sequence 298 AA;
 Query Match 5.9%; Score 19; DB 6; Length 298;
 Best Local Similarity 100.0%; Pred. No. 1.6e-09; Mismatches 0; Indels 0; Gaps 0;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 145 ALEKGLPDPVLYAEKFTP 163
 Db 100 ALEKGLPDPVLYAEKFTP 118
 RESULT 12
 ID AAB66071 standard; protein; 343 AA.
 XX
 AC AAB66071;
 XX
 XX 30-MAR-2001 (first entry)
 DT
 XX Human INTERCEPT 296.
 DE
 XX
 KW TANGO protein; INTERCEPT protein; neurological disorder;
 KW central nervous system; focal brain disorder; bipolar affective disorder;
 KW global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;
 KW senile dementia; Huntington's disease; amyotrophic lateral sclerosis;
 KW Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;
 KW neuropsychiatric; psychoactive substance use; anxiety.
 XX
 XX Homo sapiens.
 OS
 XX
 XX MO200077239-A2.
 XX
 XX 21-DEC-2000.
 PD
 XX 24-MAY-2000; 2000MO-US014858.
 PF

```

XX PA (MILL-) MILLENNIUM PHARM INC.
XX XX 14-JUN-1999; 99US-00333159.
XX PI (MILL-) MILLENNIUM PHARM INC.
XX PI McCarthy SA, Fraser CC, Sharp JD, Barnes TM;
XX DR WPI: 2001-032313/04.
XX DR N-PSDB; AAF45133, AAF45134.
XX PT TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for
XX PT screening assays and diagnostic assays and for the treatment of
XX PT neurological diseases such as Alzheimer's, Parkinson's and Huntington's
XX PT disease.
XX PS Claim 8; Fig 7; 359pp; English.
XX CC The present invention relates to TANGO or INTERCEPT proteins and coding
XX CC sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66031-B66057,
XX CC AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding
XX CC sequences are useful for the treatment of neurological disorders such as
XX CC central nervous system (CNS) disorders, CNS-related disorders, focal
XX CC brain disorders, global-diffuse cerebral disorders and other neurological
XX CC and cerebrovascular disorders. The CNS disorders include Alzheimer's
XX CC disease, senile dementia, Huntington's disease, amyotrophic lateral
XX CC sclerosis, Parkinson's, Gilles de la Tourette's syndrome, autonomic
XX CC function disorders such as hypertension and sleep disorders,
XX CC neuropsychiatric disorders, psychoactive substance use disorders,
XX CC anxiety, and bipolar affective disorder
XX SQ Sequence 343 AA;

Query Match 5.9%; Score 19; DB 4; Length 343;
Best Local Similarity 100.0%; Pred. No. 1.8e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 ALEKGLDPVLYIAEKFTP 163
DB 145 ALEKGLDPVLYIAEKFTP 163

RESULT 13
ABU08370
ID ABU08370 standard; protein; 343 AA.
XX AC ABU08370;
XX DT 03-JUN-2003 (first entry)
XX DE Amino acid sequence for human INTERCEPT 296.
XX XX
XX Human; INTERCEPT 296; TANGO; cellular process; tissue typing;
XX forensic biology; cellular protease activity; cell interaction;
XX development; blood disorder; haematopoietic cell-related disorder;
XX growth; cell proliferation; cell differentiation; gamma delta T-cell;
XX immune system; metabolic disorder; homeostatic disorder; anaemia;
XX developmental bone disorder; osteoporosis; bacterial infection;
XX thrombocytopenia; renal failure; liver disease; cystic fibrosis;
XX obesity; cancer; bronchitis; asthma; emphysema; pulmonary oedema;
XX respiratory distress syndrome; osteopathy; antibacterial; antineoplastic;
XX thrombolytic; nephrotropic; antiobesity; hepatotropic; cytoskeletal;
XX antiinflammatory; antiasthmatic.
XX XX
XX Homo sapiens.
XX OS
XX US2002182675-A1.
XX PN
XX 05-DEC-2002.
XX PD
XX 25-OCT-2001; 2001US-00042431.
XX PF
XX 14-JUN-1999; 99US-00333159.
XX PR
XX 24-MAY-2000; 2000US-00578063.

```

```

XX PA (MILL-) MILLENNIUM PHARM INC.
XX XX 14-JUN-1999; 99US-00333159.
XX PI (MILL-) MILLENNIUM PHARM INC.
XX PI McCarthy SA, Fraser CC, Sharp JD, Barnes TM;
XX DR WPI: 2001-032313/04.
XX DR N-PSDB; AAF45133, AAF45134.
XX PT TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for
XX PT screening assays and diagnostic assays and for the treatment of
XX PT neurological diseases such as Alzheimer's, Parkinson's and Huntington's
XX PT disease.
XX PS Claim 8; Fig 7; 359pp; English.
XX CC The present invention relates to TANGO or INTERCEPT proteins and coding
XX CC sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66031-B66057,
XX CC AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding
XX CC sequences are useful for the treatment of neurological disorders such as
XX CC central nervous system (CNS) disorders, CNS-related disorders, focal
XX CC brain disorders, global-diffuse cerebral disorders and other neurological
XX CC and cerebrovascular disorders. The CNS disorders include Alzheimer's
XX CC disease, senile dementia, Huntington's disease, amyotrophic lateral
XX CC sclerosis, Parkinson's, Gilles de la Tourette's syndrome, autonomic
XX CC function disorders such as hypertension and sleep disorders,
XX CC neuropsychiatric disorders, psychoactive substance use disorders,
XX CC anxiety, and bipolar affective disorder
XX SQ Sequence 343 AA;

Query Match 5.9%; Score 19; DB 6; Length 343;
Best Local Similarity 100.0%; Pred. No. 1.8e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 ALEKGLDPVLYIAEKFTP 163
DB 145 ALEKGLDPVLYIAEKFTP 163

RESULT 14
ABO32725
ID ABO32725 standard; protein; 343 AA.
XX AC ABO32725;
XX DT 17-SEP-2003 (first entry)
XX DE Secreted polypeptide-related protein #100.
XX XX
XX Human; TANGO; INTERCEPT; secreted polypeptide; immune disorder;
XX hormonal disorder; proliferative disorder; cancer; thyroid disorder;
XX diabetes; multiple sclerosis; lupus; neurological disorder; anaemia;
XX Alzheimer's disease; Parkinson's disease; cardiovascular disorder;
XX myocardial infarction; congestive heart disease; blood platelet disorder;
XX thrombocytopenia; blood vessel; atherosclerosis; vasculitis.
XX XX
XX Homo sapiens.
XX OS
XX US2003022279-A1.

```

```

XX 30-JAN-2003.
PD
XX
XX 12-JAN-2001; 2001US-00759130.
PF
XX
XX 14-JUN-1999; 99US-00333159.
PR
XX 29-JUN-1999; 99US-00342364.
PR
XX 10-SEP-1999; 99US-00393996.
PR
XX 19-OCT-1999; 99US-00420707.
PR
XX 07-JAN-2000; 2000US-00479249.
PR
XX 27-APR-2000; 2000US-00559497.
PR
XX 24-MAY-2000; 2000US-00578063.
PR
XX 16-JUN-2000; 2000US-00596194.
PR
XX 23-JUN-2000; 2000US-00602871.
PR
XX 30-JUN-2000; 2000US-00608452.
XX
XX (FRAS/) FRASER C C.
PA (BARN/) BARNES T M.
PA (SHAR/) SHARP J D.
PA (KIRST/) KIRST S J.
PA (MYER/) MYERS P S.
PA (LEIB/) LEIBY K R.
PA (HOLT/) HOLTZMAN D A.
PA (MCCA/) MCCARTHY S A.
PA (WRIK/) WRIGHTON N.
PA (MACK/) MACKAY C R.
PA (GOOD/) GOODEARL A D J.
XX
XX Fraser CC, Barnes TM, Sharp JD, Kirst SJ, Myers PS, Leiby KR;
PI Holtzman DA, Mccarthy SA, Wrighton N, Mackay CR, Goodearl ADJ;
PI
XX WPI; 2003-456290/43.
XX
XX N-PSDB; ACD66786; ACD66787.
DR
XX
XX New nucleic acid molecule encoding a secreted protein (e.g. TAMCO 202,
PT TAMCO 210 or INTERCEPT 217), useful for diagnosing, preventing or
PT treating disorders such as cancer, diabetes or atherosclerosis, and in
PT forensic biology.
XX
XX Claim 9; Fig 30A-30C; 482pp; English.
XX
XX The invention relates to secreted polypeptide-related proteins and
XX nucleic acids (TAMCO and INTERCEPT proteins and nucleic acids). The
XX nucleic acids, proteins and antibodies specific to the proteins are
XX useful in screening assays, predictive medicine (e.g. diagnostic assays,
XX prognostic assays, monitoring clinical trials and pharmacogenetics) and
XX prophylactic and therapeutic methods. The sequences are used in
XX diagnosing, preventing or treating proliferative disorders (e.g.
XX cancers), hormonal disorders (e.g. diabetes or thyroid disorders), immune
XX disorders (e.g. multiple sclerosis or lupus), neurological disorders
XX (e.g. Alzheimer's disease or Parkinson's disease), cardiovascular
XX disorders (e.g. myocardial infarction or congestive heart disease), blood
XX platelet disorders (e.g. thrombocytopenia or anaemia) and disorders
XX involving blood vessels (e.g. atherosclerosis or vasculitis). The nucleic
XX acids may also be used in chromosome mapping, tissue typing and forensic
XX biology, and as surrogate markers. This sequence represents a secreted
XX polypeptide-related protein of the invention. Note: The sequence data for
XX this patent was obtained in electronic format directly from USPTO at
XX seqdata.uspto.gov/sequence.html
XX
XX Sequence 343 AA;
SQ
XX
XX Query Match 5.9%; Score 19; DB 6; Length 343;
XX Best Local Similarity 100.0%; Pred. No. 1.8e-09;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 145 ALEKGLDPVLYLAEKFTP 163
DB 145 ALEKGLDPVLYLAEKFTP 163

```

```

ID ABG03520 standard; protein; 445 AA.
XX
XX AC ABG03520;
XX
XX 13-FEB-2002 (first entry)
DT
XX
XX Novel human diagnostic protein #3511.
DE
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
XX
XX WO200175067-A2.
FN
XX
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001MO-US008631.
PF
XX
XX 31-MAR-2000; 2000US-00540217.
XX
XX 23-AUG-2000; 2000US-00649167.
PR
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX
XX WPI; 2001-639362/73.
XX
XX N-PSDB; AAS67707.
DR
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
PT
XX
XX Claim 20; SEQ ID NO 33879; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
XX amino acid sequences of the invention. Note: The sequence data for this
XX patent did not appear in the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pat_sequences
XX
XX Sequence 445 AA;
SQ
XX
XX Query Match 5.9%; Score 19; DB 4; Length 445;
XX Best Local Similarity 100.0%; Pred. No. 2.3e-09;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 145 ALEKGLDPVLYLAEKFTP 163
DB 145 ALEKGLDPVLYLAEKFTP 163

```

Search completed: February 23, 2004, 19:33:18
Job time : 60 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 21:58:00 ; Search time 5685.97 Seconds

(without alignments) 10984.465 Million cell updates/sec

Title: US-09-936-456-1

Perfect score: 1441

Sequence: 1 aaagtaacgctacagacag.....aaaaaaaaaaaaaaaa 1441

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 2167151695 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Database :

```
1: gb_ba:*
2: gb_hg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_hg_hum:*
31: em_hg_inv:*
32: em_hg_other:*
33: em_hg_mus:*
34: em_hg_pln:*
35: em_hg_rod:*
36: em_hg_man:*
37: em_hg_vrt:*
38: em_sy:*
39: em_hggo_hum:*
40: em_hggo_mus:*
41: em_hggo_other:*
```

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Match	Length	DB ID	Description
1	1441	100.0	1441	6 AX035346	AX035346 Sequence
2	976	67.7	1740	6 HSM806241	BX537581 Homo sapi
3	425	29.5	498	6 AX035348	AX035348 Sequence
4	405	28.1	166937	2 AC012255	AX012255 Homo sapi
5	405	28.1	181312	6 AC091117	AC091117 Homo sapi
6	392	27.2	2684	6 AX714511	AX714511 Sequence
7	392	27.2	2684	6 AK056896	AK056896 Homo sapi
8	392	27.2	156534	2 AC009700	AC009700 Homo sapi
9	392	27.2	171444	2 AC087790	AC087790 Homo sapi
10	354	24.6	156534	2 AC009700	AC009700 Homo sapi
11	217	15.1	217	9 H5183H12F	Z57552 H. sapiens C
12	166	11.5	1923	9 BC029819	BC029819 Homo sapi
13	136	9.4	218	9 H5183A12R	Z59954 H. sapiens C
14	39	2.7	1725	9 BC004161	BC004161 Homo sapi
15	39	2.7	1726	9 BC003157	BC003157 Homo sapi
16	38	2.6	1260	9 BC017580	BC017580 Homo sapi
17	38	2.6	2165	6 BD275921	BD275921 Uncouplin
18	38	2.6	2384	6 AX780522	AX780522 Sequence
19	38	2.6	2384	6 AX780523	AX780523 Sequence
20	38	2.6	3993	6 BC036370	BC036370 Homo sapi
21	38	2.6	5642	6 AX345771	AX345771 Sequence
22	38	2.6	10480	6 AX347103	AX347103 Sequence
23	38	2.6	174780	2 AC134964	AC134964 Canis fam
24	38	2.6	180059	2 AC023928	AC023928 Homo sapi
25	38	2.6	180763	2 AC067826	AC067826 Homo sapi
26	38	2.6	185951	2 AC024386	AC024386 Homo sapi
27	38	2.6	186051	2 AC015683	AC015683 Homo sapi
28	37	2.6	1426	10 AF175224	AF175224 Rattus no
29	37	2.6	1640	9 BC050738	BC050738 Homo sapi
30	37	2.6	1678	9 HSM802263	AL137530 Homo sapi
31	37	2.6	1724	9 BC003101	BC003101 Homo sapi
32	37	2.6	2645	3 AY118446	AY118446 Drosophila
33	37	2.6	3734	9 HSM807535	BX647390 Homo sapi
34	37	2.6	3989	9 HSM805415	AL634162 Homo sapi
35	37	2.6	3990	9 HSM805416	AL634352 Homo sapi
36	37	2.6	135628	2 AC067764	AC067764 Homo sapi
37	37	2.6	153166	2 AC048357	AC048357 Homo sapi
38	37	2.6	190371	2 AC116830	AC116830 Mus muscu
39	37	2.6	349980	6 AX344557	AX344557 Sequence
40	36	2.5	1059	3 AF036416	AF036416 Trypanoso
41	36	2.5	1407	10 BC013450	BC013450 Mus muscu
42	36	2.5	1603	10 BC024516	BC024516 Mus muscu
43	36	2.5	1713	9 BC003624	BC003624 Homo sapi
44	36	2.5	2155	9 BC015749	BC015749 Homo sapi
45	36	2.5	2255	6 AR168373	AR168373 Sequence

ALIGNMENTS

```
RESULT 1
LOCUS AX035346 1441 bp DNA linear PAT 15-NOV-2000
DEFINITION Sequence 1 from Patent WO0053748.
ACCESSION AX035346
VERSION AX035346.1 GI:11191064
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryote; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Bruck/C.E., Vinals/Y.D., Coche/T. and Casseart/J.P.
AUTHORS Novel compounds
TITLE Patent: WO 0053748-A 1 14-SEP-2000;
JOURNAL
```

BRUCK CLAUDE ELVIRE MARIE (BE) ; SMITHKLINE BEECHAM BIOLOG (BE) ;
VINUS Y DE BASSELS CARLOTTA (BE) ; COCHE THIERRY (BE) ; CASPART
JEAN POL (BE)

FEATURES

source
1..1441
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 1441; DB 6; Length 1441;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1441; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAGTACGGCTACAGACAGATGAGAAATAGTTGCTGCGCGCTAGAAAACTCTGCG 60
DB 1 AAAGTACGGCTACAGACAGATGAGAAATAGTTGCTGCGCGCTAGAAAACTCTGCG 60
QY 61 GTACCAACCCCGAGAGCGTTGAGAGACCCCACTTCCCTTAAAGAGAGTGA 120
DB 61 GTACCAACCCCGAGAGCGTTGAGAGACCCCACTTCCCTTAAAGAGAGTGA 120
QY 121 GGACTGAGACTTCAACAGCCCACTGGTCCAGCTTGTACGAAAGAGAGCCAGAGAC 180
DB 121 GGACTGAGACTTCAACAGCCCACTGGTCCAGCTTGTACGAAAGAGAGCCAGAGAC 180
QY 181 GCGCTCTCCCGGCTCCAGGACCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
DB 181 GCGCTCTCCCGGCTCCAGGACCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
QY 241 ACTCGGCGCGCGCTGAGAGCATGACCTGTGTGAGAACGCGCTTATTTTACCCAGCCC 300
DB 241 ACTCGGCGCGCGCTGAGAGCATGACCTGTGTGAGAACGCGCTTATTTTACCCAGCCC 300
QY 301 CGGCAATGCGGAGGCTTCAAGCTGCTCATGCTTATTTAGTGTGCTGCTA 360
DB 301 CGGCAATGCGGAGGCTTCAAGCTGCTCATGCTTATTTAGTGTGCTGCTA 360
QY 361 GCAGCAAGCTTCTGCTGCTCTTTGCGGGGATCCGCTGCGCACTGCGCTGCTGCTG 420
DB 361 GCAGCAAGCTTCTGCTGCTCTTTGCGGGGATCCGCTGCGCACTGCGCTGCTGCTG 420
QY 421 GTGAGAGTCTTCTGAGTGTGCTTCAATAGGCGAGAAATTTGCTGCTGCTGCTGCA 480
DB 421 GTGAGAGTCTTCTGAGTGTGCTTCAATAGGCGAGAAATTTGCTGCTGCTGCTGCA 480
QY 481 GAATGCTTGTGAGTGTGCTGAGAACCAACATCTTCAAAAGCTTCAAGCGAGCGCGC 540
DB 481 GAATGCTTGTGAGTGTGCTGAGAACCAACATCTTCAAAAGCTTCAAGCGAGCGCGC 540
QY 541 GTTACAGCCCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
DB 541 GTTACAGCCCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
QY 601 ACCCGAGTGCATGCTGAGAACGACATTTGACTCAACAGAGCTTCACTGCGCTGCTG 660
DB 601 ACCCGAGTGCATGCTGAGAACGACATTTGACTCAACAGAGCTTCACTGCGCTGCTG 660
QY 661 AAAGAGAAATTTACGCGCGGAGTACGAGAACGACATGAGAGGCGCTGCGGACCCAGTG 720
DB 661 AAAGAGAAATTTACGCGCGGAGTACGAGAACGACATGAGAGGCGCTGCGGACCCAGTG 720
QY 721 CTCTACCTGCGGAGAGTTCACACGAGTAGCCCTTGTGCGCTGTACACACAGTACAC 780
DB 721 CTCTACCTGCGGAGAGTTCACACGAGTAGCCCTTGTGCGCTGTACACACAGTACAC 780
QY 781 CTGCGGAGACATGAGCGCTGCGGAGACGATGAGTGTGCTGCTGCTGCTGCTGCTGCTG 840
DB 781 CTGCGGAGACATGAGCGCTGCGGAGACGATGAGTGTGCTGCTGCTGCTGCTGCTGCTG 840
QY 841 AACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
DB 841 AACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900

QY 901 GCCTTCGGCTCTCCGAGGCTCTTCCGCTTGGCTTCATCTAGAGTGGCGCTGCGCG 960
DB 901 GCCTTCGGCTCTCCGAGGCTCTTCCGCTTGGCTTCATCTAGAGTGGCGCTGCGCG 960
QY 961 CTCCGCTTGGCTCTCCGAGGCTCTTCCGCTTGGCTTCATCTAGAGTGGCGCTGCGCG 1020
DB 961 CTCCGCTTGGCTCTCCGAGGCTCTTCCGCTTGGCTTCATCTAGAGTGGCGCTGCGCG 1020
QY 1021 GCAACCGGCTCTCCGAGGCTCTTCCGCTTGGCTTCATCTAGAGTGGCGCTGCGCG 1080
DB 1021 GCAACCGGCTCTCCGAGGCTCTTCCGCTTGGCTTCATCTAGAGTGGCGCTGCGCG 1080
QY 1081 CCGAGCGCTCTTCCGAGGCTCTTCCGCTTGGCTTCATCTAGAGTGGCGCTGCGCG 1140
DB 1081 CCGAGCGCTCTTCCGAGGCTCTTCCGCTTGGCTTCATCTAGAGTGGCGCTGCGCG 1140
QY 1141 GAGTCACTCTTATCTTCCGAGGCTCTTCCGCTTGGCTTCATCTAGAGTGGCGCTGCGCG 1200
DB 1141 GAGTCACTCTTATCTTCCGAGGCTCTTCCGCTTGGCTTCATCTAGAGTGGCGCTGCGCG 1200
QY 1201 TGATCACTCTTATCTTCCGAGGCTCTTCCGCTTGGCTTCATCTAGAGTGGCGCTGCGCG 1260
DB 1201 TGATCACTCTTATCTTCCGAGGCTCTTCCGCTTGGCTTCATCTAGAGTGGCGCTGCGCG 1260
QY 1261 CGAGGCTCTTATCTTCCGAGGCTCTTCCGCTTGGCTTCATCTAGAGTGGCGCTGCGCG 1320
DB 1261 CGAGGCTCTTATCTTCCGAGGCTCTTCCGCTTGGCTTCATCTAGAGTGGCGCTGCGCG 1320
QY 1321 AGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
DB 1321 AGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
QY 1381 TCTGTAATTAACCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1440
DB 1381 TCTGTAATTAACCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1440
QY 1441 A 1441
DB 1441 A 1441

RESULT 2

HSMB06241

LOCUS HSMB06241 1740 bp mRNA linear PRI 17-JUN-2003

DEFINITION Homo sapiens mRNA; cDNA DKFP686C04213 (from clone DKFP686C04213).

ACCESSION BX537581

VERSION BX537581.1 GI:31873659

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (17-JUN-2003) MIPS, Ingolstaedter Landstr.1, D-85764
Neuberg, GERMANY
Molecular Genome Analysis, German Cancer
Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de,
Braunschweig/Germany) within the cDNA sequencing consortium of the
German Genome Project.
This clone (DKFP686C04213) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de Further
information about the clone and the sequencing project is available
at http://mips.gsf.de/proj/cDNA/.

FEATURES

source
1..1740
/organism="Homo sapiens"
/mol_type="mRNA"


```

/db_xref="taxon:9606"
/map_115g15"
/clone="DK2p686CC4213"
/tissue_type="human rectum tumor"
/clone_lib="686 (synonym: h1cc3). Vector pSport1-5'1; host
DH10B." sites 5f11A + 5f11B"
/dev_stage="adult"
polyA_signal
polyA_site
1694_..1699
1720

```

```

Query Match      67.7%; Score 976; DB 9; Length 1740;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1026; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY	405	GGCGCTGGTTTGGTTGGTGAGAGTCTCTTCACGTCTCTTCATATGAGGCGAGAAATGTGTGGC	464
Db	712	GGCGTGGTTTTGGTTGGTGTAGAGTCTCTTCACGTCTCTTCATATGAGGCGAGAAATGTGTGGC	771
QY	465	TGTGCACTTCAGTGACGAAATGGTTGTGTGGGTACATGATMACCAACACACATCTCTACAAAGC	524
Db	772	TGTGCACTTCAGTGACGAAATGGTTGTGTGGGTACATGATMACCAACACATCTCTACAAAGC	831
QY	525	CTTCAGGCGACGCCGGCTTACAGCCCCGTGTGGTCTGCTGTGGGCTCTGAGGGCATTTA	584
Db	832	CTTCAGGCGACGCCGGCTTACAGCCCCGTGTGGTCTGCTGTGGGCTCTGAGGGCATTTA	891
QY	585	TATTACACTACAGGGAGCCCGAGTGCATCAGCTCAGACTAGAACCATTTGACTACACAGACA	644
Db	892	TATTACACTACAGGGAGCCCGAGTGCATCAGCTCAGACTAGAACCATTTGACTACACAGACA	951
QY	645	GTTTCACCTGGCGCTCTGAAAAGAAATTACGCCCGCGAGTACGCCGATCGCATCTGAGAGGG	704
Db	952	GTTTCACCTGGCGCTCTGAAAAGAAATTACGCCCGCGAGTACGCCGATCGCATCTGAGAGGG	1011
QY	705	GCATCGGAGACCCAGATGCTCTTACCTCTGACGAGAGAGTTTACACCGAGTAGCCCTTGCGGCT	764
Db	1012	GCATCGGAGACCCAGATGCTCTTACCTCTGACGAGAGAGTTTACACCGAGTAGCCCTTGCGGCT	1071
QY	765	GTACACACAGTACCACTTGCGGAGAACATAAGCCTCGGCGACCGCTATGGGTGGCGTTCTG	824
Db	1072	GTACACACAGTACCACTTGCGGAGAACATAAGCCTCGGCGACCGCTATGGGTGGCGTTCTG	1131
QY	825	CTTCTGGCTCTCTCCACAGGTGCTGTCCACACGCGCGAGCCCGCGCTACAGAGAGCTGTGC	884
Db	1132	CTTCTGGCTCTCTCTCCACAGGTGCTGTCCACACGCGCGAGCCCGCGCTACAGAGAGCTGTGC	1191
QY	885	ACTGTGACCAACCGAGAGCTTTGCGCGCTCTTGCGGGTCTTGCGCTTGACTCTCATCTCTAG	944
Db	1192	ACTGTGACCAACCGAGAGCTTTGCGCGCTCTTGCGGGTCTTGCGCTTGACTCTCATCTCTAG	1251
QY	945	CGATCGGCTGTGCACGCGCTCGAGCTTAAAGCTCCCGCGCTCACACCTCAGTACGAGCGCGCG	1004
Db	1252	CGATCGGCTGTGCACGCGCTCGAGCTTAAAGCTCCCGCGCTCACACCTCAGTACGAGCGCGCG	1311
QY	1005	CTTCTGGGTACGCTGAGCAACCGGCGCTCTGTGCTCTTCTCTGGAAGGGCGCTGTGTAG	1064
Db	1312	CTTCTGGGTACGCTGAGCAACCGGCGCTCTGTGCTCTTCTCTGGAAGGGCGCTGTGTAG	1371
QY	1065	TCTCCAGTATGTTGGGCCAGCGGCTTTGGACCCCTTGAGACCAAGAGGCGCACAGGACATG	1122
Db	1372	TCTCCAGTATGTTGGGCCAGCGGCTTTGGACCCCTTTGAGACCAAGAGGCGCACAGGACATG	1431
QY	1125	CAGCCAGAGAGAGGGGGCTCACCTCTTATCTCTGGGACCCACTGTGCACAGCAGGCGCG	1184
Db	1432	CAGCCAGAGAGAGGGGGCTCACCTCTTATCTCTGGGACCCACTGTGCACAGCAGGCGCG	1491
QY	1185	TCTCCAGACCTTAAATGTATATACCAACTTAACTGTGAGGGGGAGCCCAATGTGGACTCTCT	1244
Db	1492	TCTCCAGACCTTAAATGTATATACCAACTTAACTGTGAGGGGGAGCCCAATGTGGACTCTCT	1551
QY	1245	CCCCGCTTGGGACATTCGAGGCGGGAGACGTGCCCGCGAGGCTGTGGCCACGAGAGAC	1304

Db	1552	CCCCGCTTGGGACATCGCAGAGCCGGGAAACACTGTGCCCGCCAGAGCCTGGGCAGAGAGGC	1611
QY	1305	TCGAGGAGAGGCACTGAAGCGCTGTGGCGCGAGAGCCTCGGAACATCCGACAGGACACGAGGA	1364
Db	1612	TCGAGGAGAGGCACTGAAGCGCTGTGGCGCGAGAGCCTCGGAACATCCGACAGGACACGAGGA	1671
QY	1365	AAATCTCCCTGGGGCGCATCTGTAAATAAACCCTTTTCTTTGTTTTTAAAAAAAAA	1424
Db	1672	AAATCTCCCTGGGGCGCATCTGTAAATAAACCCTTTTCTTTGTTTTTAAAAAAAAA	1731
QY	1425	AAAAAAA 1431	
Db	1732	AAAAAAA 1738	

RESULT 3	AX035348	498 bp	DNA	linear	PAT 15-NOV-2000
LOCUS	AX035348				
DEFINITION	Sequence 3 from Patent WO0053748.				
ACCESSION	AX035348				
VERSION	AX035348.1	GI:1191065			
KEYWORDS					
SOURCE					
ORGANISM	Homo sapiens (human)				
REFERENCE	Homo sapiens				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.				
JOURNAL	1				
	Bruck,C.E., Vinalis,Y.D., Coche,T. and Casart,J.P.				
	Novel compounds				
	Patent: WO 0053748-A 3 14-SEP-2000;				
	BRUCK CLAUDE ELVINE MARIE (BE) ; SMITHKLINE BEECHAM BIOLOG (BE) ;				
	VINALIS Y DE BASSELO CARLOTA (BE) ; COCHE THIERRY (BE) ; CASART				
	JEAN POL (BE)				
FEATURES	Location/Qualifiers				
source	1..498				
	/organism="Homo sapiens"				
	/mol_type="unassigned DNA"				
	/db_xref="taxon:9606"				
ORIGIN					
	Query Match 29.5%; Score 425; DB 6; Length 498;				
	Best Local Similarity 100.0%; Prid. No. 5.8e-224;				
	Matches 425; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	998 GGGCGCCCTTCCTGGGTCACGCTGGCAACCGGCGCTGTCGCTCTTCCTGGAGGGGCG	105			
DB	62 GCGCGCGCTTCGGGTCACGCTGGCAACCGGCGCTTCCTTCCTTCCTGGAGGGGCG	121			
QY	1058 TGGTGAAGTCTCAGTATGTTGGGCCACGCGCTTCCTGGACCTTCCTGACCAAGCGCA	1117			
DB	122 TGGTGAAGTCTCAGTATGTTGGGCCACGCGCTTCCTGGACCTTCCTGACCAAGCGCA	181			
QY	1118 AGGACGTGACGACGAGAGAGAGGGGGCTCACTCTTATCCCGGCGACCACTGACCAAGC	1177			
DB	182 AGGACGTGACGACGAGAGAGGGGGCTCACTCTTATCTCGGACCACTGACCAAGC	241			
QY	1178 AGGCGCGCTCTCCAGACTTAAATGTATCACTAACTGTGAGGGGGAGCCCAATCTGG	1237			
DB	242 AGGCGCGCTCTCCAGACTTAAATGTATCACTAACTGTGAGGGGGAGCCCAATCTGG	301			
QY	1238 ACTCTCTTCCCGCTTGGGACATGTGACGCGGGGAACATGTGCGGCGCAAGGCTGGGGCA	1297			
DB	302 ACTCTCTTCCCGCTTGGGACATGTGACGCGGGGAACATGTGCGGCGCAAGGCTGGGGCA	361			
QY	1298 GGAAGCTCCAGAAAGGACACTGACGCTGTGCGCGAGGCTTCGACATCCGAGGCA	1357			
DB	362 GGAAGCTCCAGAAAGGACACTGACGCTGTGCGCGAGGCTTCGACATCCGAGGCA	421			
QY	1358 CCAAGGAAAGTCTCTGGGGGCACTGTGAAATAAAGCTTTTCTTTGTTTAAAA	1417			
DB	422 CCAAGGAAAGTCTCTGGGGGCACTGTGAAATAAAGCTTTTCTTTGTTTAAAA	481			
QY	1418 AAAAA 1422				

```

|||||
Db      482 AAAA 486

RESULT 4
AC012255
LOCUS
DEFINITION Homo sapiens chromosome 15 clone RP11-109D20 map 15, WORKING DRAFT
SEQUENCE, 22 unordered pieces.
AC012255      166937 bp    DNA      linear      HTG 06-MAY-2001
AC012255.4    GI:7684440
VERSION
AC012255.4    HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE
Homo sapiens
Homo sapiens (human)
ORGANISM
Mammalia; Euteleostomi; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 166937)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-109D20
Unpublished
2 (bases 1 to 166937)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckert,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeRellano,K., Dewar,K., Domingo,M., Donegan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Horton,L.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Holland,J.C., Johnson,R., Jones,C., Kahn,L., Karatas,A., Klein,J.,
Lehoczky,J., Lien,C., Locke,K., Macdonald,P., Margulis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meidrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Sudramanian,A., Talamas,J.,
Testaye,S., Tittell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,W.
Direct Submission
Submitted (21-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Smit,A.F.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3737
Center clone name: 109.D.20
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 153800 bases at least Q40
Consensus quality: 160606 bases at least Q30
Consensus quality: 163275 bases at least Q20
Insert size: 164837; sum-of-coverage
Quality coverage: 4.0 in Q20 bases; sum-of-coverage
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2087: contig of 2087 bp in length
* 2088 2187: gap of 100 bp
* 2188 3626: contig of 1439 bp in length
* 3627 3727: gap of 100 bp
* 3727 4845: contig of 1119 bp in length

```

```

FEATURES
source
*
* 4846 4945: gap of 100 bp
* 4946 6702: contig of 1757 bp in length
* 6703 6802: gap of 100 bp
* 6803 8325: contig of 1524 bp in length
* 8326 8427: gap of 100 bp
* 8428 10494: contig of 2068 bp in length
* 10495 10595: gap of 100 bp
* 10596 13215: contig of 2621 bp in length
* 13216 13316: gap of 100 bp
* 13317 16452: contig of 3136 bp in length
* 16453 16552: gap of 100 bp
* 16553 20843: contig of 4292 bp in length
* 20844 20944: gap of 100 bp
* 20945 24941: contig of 3997 bp in length
* 24942 25041: gap of 100 bp
* 25042 29902: contig of 4861 bp in length
* 29903 30002: gap of 100 bp
* 30003 34189: contig of 4188 bp in length
* 34190 34289: gap of 100 bp
* 34290 38387: gap of 4097 bp in length
* 38388 45827: contig of 7341 bp in length
* 45828 45928: gap of 100 bp
* 45929 51823: contig of 5802 bp in length
* 51824 51830: gap of 100 bp
* 51831 60123: contig of 8294 bp in length
* 60124 60224: gap of 100 bp
* 60225 73533: contig of 13309 bp in length
* 73534 73633: gap of 100 bp
* 73634 87433: contig of 13801 bp in length
* 87434 87534: gap of 100 bp
* 87535 101708: contig of 14175 bp in length
* 101709 101808: gap of 100 bp
* 101809 121193: contig of 19385 bp in length
* 121194 121294: gap of 100 bp
* 121295 139051: contig of 17788 bp in length
* 139052 139152: gap of 100 bp
* 139153 166937: contig of 27766 bp in length.
Location/Qualifiers
1..166937
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-109D20"
/clone_id="RP11-11 Human Male BAC"
1..2087
/note="assembly_fragment"
2188..3626
/note="assembly_fragment"
3727..4845
/note="assembly_fragment"
4946..6702
/note="assembly_fragment"
6803..8326
/note="assembly_fragment"
8427..10494
/note="assembly_fragment"
10595..13215
/note="assembly_fragment"
13316..16452
/note="assembly_fragment"
16553..20843
/note="assembly_fragment"
20944..24941
/note="assembly_fragment"
25042..29902
/note="assembly_fragment"
30003..34189
/note="assembly_fragment"
34290..38387
/note="assembly_fragment"

```

```

misc_feature      38487..45827
                    /note="assembly_fragment"
misc_feature      45928..51729
                    /note="assembly_fragment"
misc_feature      51830..60123
                    /note="assembly_fragment"
misc_feature      60224..73532
                    /note="assembly_fragment"
misc_feature      73633..87433
                    /note="assembly_fragment"
                    clone_end:SP6
                    vector_side:right"
misc_feature      87534..101708
                    /note="assembly_fragment"
misc_feature      101809..121193
                    /note="assembly_fragment"
misc_feature      121294..139051
                    /note="assembly_fragment"
misc_feature      139152..166937
                    /note="assembly_fragment"
                    clone_end:T7
                    vector_side:right"

ORIGIN
Query Match      28.1%; Score 405; DB 2; Length 166937;
Best Local Similarity 100.0%; Pred. No. 7.4e-213;
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  AAAGTAAAGGCTACAGACAGAGAAATAGTTTCGCTCGCCGGCTGAGAAAACCTCTGTG 60
DB      157716  AAAGTAAAGGCTACAGACAGAGAAATAGTTTCGCTCGCCGGCTGAGAAAACCTCTGTG 157775
QY      61  GTACCAACCCCGAGAGCGTTGAGACAGACCCACCTCCACGCTTCCTTACGAGAGGTGCA 120
DB      157776  GTACCAACCCCGAGAGCGTTGAGACAGACCCACCTCCACGCTTCCTTACGAGAGGTGCA 157835
QY      121  GGACTCAGACTTTCACAGACCCACCTCGCTCCCAAGCTTGTACGAAAGAGACCGCAAGAC 180
DB      157836  GGACTCAGACTTTCACAGACCCACCTCGCTCCCAAGCTTGTACGAAAGAGACCGCAAGAC 157895
QY      181  GCGCTCTCCCGCGTCCAGAGACCCCGAGCTTGCTGCTGCTCCGCCGCTGCGTGCAGC 240
DB      157896  GCGCTCTCCCGCGTCCAGAGACCCCGAGCTTGCTGCTGCTCCGCCGCTGCGTGCAGC 157955
QY      241  ACTGGCGCGGCGTGCAGATACCTCTGGAAAGCGCGTACTGCTTTTAAACCCCGACCC 300
DB      157956  ACTGGCGCGGCGTGCAGATACCTCTGGAAAGCGCGTACTGCTTTTAAACCCCGACCC 158015
QY      301  CGGCATGCGCAGAGCTTCAGAGCTTCCACTGCTCATCTTATCTAGTGTTTTGCTCTA 360
DB      158016  CGGCATGCGCAGAGCTTCAGAGCTTCCACTGCTCATCTTATCTAGTGTTTTGCTCTA 158075
QY      361  GCAGCAGCTTCTGCTCATCTTCCCGGGGATCCGTGGCCACTCG 405
DB      158076  GCAGCAGCTTCTGCTCATCTTCCCGGGGATCCGTGGCCACTCG 158120

RESULT 5
AC091117 181312 bp DNA linear PRI 20-FEB-2002
LOCUS Homo sapiens chromosome 15 clone RP11-109D20 map 15q15, complete
DEFINITION sequence.
ACCESSION AC091117
VERSION AC091117.5 GI:18767395
KEYWORDS HTG
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 181312)
AUTHORS Burke,J., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
Pate,D., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
Pate,D. and Hood,L.

```

```

TITLE
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 181312)
AUTHORS Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
Pate,D. and Hood,L.
TITLE Direct Submission
JOURNAL Submitted (30-MAR-2001) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA
REFERENCE 3 (bases 1 to 181312)
AUTHORS Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
Pate,D. and Hood,L.
TITLE Direct Submission
JOURNAL Submitted (13-FEB-2002) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA
REFERENCE 4 (bases 1 to 181312)
AUTHORS Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
Pate,D. and Hood,L.
TITLE Direct Submission
JOURNAL Submitted (14-FEB-2002) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA
REFERENCE 5 (bases 1 to 181312)
AUTHORS Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
Pate,D. and Hood,L.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-2002) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA
COMMENT On Feb 20, 2002 this sequence version replaced gi:18653547.
----- Genome Center
Center: Multimegabase Sequencing Center
Center code: UMSC
Web site: http://chroma.mbr.washington.edu/msg_www
Contact: leewen@systemsbiology.org
----- Summary Statistics
Sequencing vector: pUC18; 1.08752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-primer Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399

Note: Data from overlapping clone AC090888 [Drafting center: WMBR]
was added for finishing. RP11-109D20 spans bases 1-174236 of this
sequence. Bases 174237-179373 derive from clone RP11-276K9,
AC087790 [Drafting center: WMBR] to establish the overlap with the
next BAC in our tiling path. Bases 179373-181312 derive from
finished clone CTD-2651B20, AC051619 [Drafting center: UMSC] in
order to give overlap required for determining long range
contiguity.

FEATURES
source
location/Qualifiers
1..181312
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q15"
/clone="RP11-109D20"
/clone_lib="RP11 human BAC library 11"
/note="This clone overlaps CTD-2014N11 AC090888 and
RP11-276K9 AC087790. Data from overlapping BACs were added
and the consensus sequence determined from RP11-109D20 to
the extent possible."
1..89137
/note="Overlap with CTD-2014N11, AC090888"
36107..36120
/note="low quality data"
40470..40820
misc_feature

```

misc_feature 53030..53480 /note="sequence data generated from subcloned PCR product"
 /note="sequence data generated from subcloned PCR product"
 misc_feature 86982..86983 /note="low quality data"
 misc_feature 95610..96030 /note="sequence data generated from subcloned PCR product"
 misc_feature 96830..97420 /note="sequence data generated from subcloned PCR product"
 misc_feature 98260..98420 /note="single clone coverage"
 misc_feature 98850..98852 /note="low quality data"
 misc_feature 100981..101003 /note="low quality data"
 misc_feature 174237..179372 /note="overlap with RPI1-276K9, AC087790. This sequence data was added to establish the overlap between the BACs in the tiling path"
 misc_feature 179373..181312 /note="overlap with CTD-2651B20, AC051619. This data is added to give overlap for long range contiguity."

ORIGIN

Query Match 28.1%; Score 405; DB 9; Length 181312;
 Best Local Similarity 100.0%; Fred. No. 7.4e-213; Indels 0; Gaps 0;
 Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAGTAACGGCTACAGACAGTGAAGAAATAGTTCCGCTCGCGCTAGAAAACTCTGCG 60
 DB 165016 AAAGTAACGGCTACAGACAGTGAAGAAATAGTTCCGCTCGCGCTAGAAAACTCTGCG 165075
 QY 61 GTACCAACCCAGAGGCTTGAAGAGACCCCACTCCAGCTTCTTAAAGAGAGTGA 120
 DB 165076 GTACCAACCCAGAGGCTTGAAGAGACCCCACTCCAGCTTCTTAAAGAGAGTGA 165135
 QY 121 GGACTCAGACTTACAGAGCCCACTCGATCCAGCTTGTAGCGAAGAGAGCCAGAGAC 180
 DB 165136 GGACTCAGACTTACAGAGCCCACTCGATCCAGCTTGTAGCGAAGAGAGCCAGAGAC 165195
 QY 181 GCGCTTCCCGCGCTCCAGAGAGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
 DB 165196 GCGCTTCCCGCGCTCCAGAGAGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 165255
 QY 241 ACTCGGCGCGCTGAGACAGTGAAGAAATAGTTCCGCTCGCGCTAGAAAACTCTGCG 300
 DB 165256 ACTCGGCGCGCTGAGACAGTGAAGAAATAGTTCCGCTCGCGCTAGAAAACTCTGCG 165315
 QY 301 CGGCAATGCGCGAGGCTTCAAGCGTTCACCTGCTCATCTGTTATTTGAGCTCTA 360
 DB 165316 CGGCAATGCGCGAGGCTTCAAGCGTTCACCTGCTCATCTGTTATTTGAGCTCTA 165375
 QY 361 GCAGCAAGCTTCTGCTCATCTGCGCGGAGATCCGCGGCACTGCG 405
 DB 165376 GCAGCAAGCTTCTGCTCATCTGCGCGGAGATCCGCGGCACTGCG 165420

RESULT 6
 AX714511/c 2684 bp DNA linear PAT 15-APR-2003
 LOCUS AX714511
 DEFINITION Sequence 1195 from Patent EP1293569.
 ACCESSION AX714511
 VERSION AX714511.1 GI:29889464
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE
 AUTHORS Isegai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S.,
 Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R.,
 Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and
 Masuno, Y.

TITLE Full-length cDNAs
 JOURNAL Patent: EP 1293569-A 1195 19-MAR-2003;
 Helix Research Institute (JRI); Research Association for
 Biotechnology (JRI)
 FEATURES
 source
 Location/Qualifiers
 1..2684
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN

Query Match 27.2%; Score 392; DB 6; Length 2684;
 Best Local Similarity 100.0%; Fred. No. 1.2e-205; Indels 0; Gaps 0;
 Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1027 GAGGTCCTGTCCTCTTCTCTGAGAGGCGCGGTGATGCTCCAGTATGTCGCCACG 1086
 DB 2343 GAGGTCCTGTCCTCTTCTCTGAGAGGCGCGGTGATGCTCCAGTATGTCGCCACG 2284
 QY 1087 GCTCTTCCGACCCCTTCTGAGACCAAGGCCCAAGACTGCGCCAGAGAGAGGGGCTCA 1146
 DB 2283 GCTCTTCCGACCCCTTCTGAGACCAAGGCCCAAGACTGCGCCAGAGAGAGGGGCTCA 2224
 QY 1147 CCTTTATCTCGGCGACCCCACTGACAGAGGCGCTCTCCAGACTTAAATATATC 1206
 DB 2223 CCTTTATCTCGGCGACCCCACTGACAGAGGCGCTCTCCAGACTTAAATATATC 2164
 QY 1207 ACCACTAATCTGTGAGGGGAGCCCAATCTGATCTCTCCCGCTTGGAGCATCGCAG 1266
 DB 2163 ACCACTAATCTGTGAGGGGAGCCCAATCTGATCTCTCCCGCTTGGAGCATCGCAG 2104
 QY 1267 CCGGAGACAGTGCCTCGGAGGCTGAGGCGCGAGAGAGCTCAGAGAGGAGCACTAGCGCT 1326
 DB 2103 CCGGAGACAGTGCCTCGGAGGCTGAGGCGCGAGAGAGCTCAGAGAGGAGCACTAGCGCT 2044
 QY 1327 GCTGCGGAGGCTCGGAGCATCGGAGGACACAGGAAAGTCTCTGGGCGATCTGTA 1386
 DB 2043 GCTGCGGAGGCTCGGAGCATCGGAGGACACAGGAAAGTCTCTGGGCGATCTGTA 1984
 QY 1387 AATAAACCTTTTCTTTGTTTAAAAA 1418
 DB 1983 AATAAACCTTTTCTTTGTTTAAAAA 1952

RESULT 7
 AK056896/c 2684 bp mRNA linear PRI 01-AUG-2002
 LOCUS AK056896
 DEFINITION Homo sapiens cDNA FLJ32334, clone PROST2005426.
 ACCESSION AK056896
 VERSION AK056896.1 GI:16552419
 KEYWORDS oligo cloning; fls (full insert sequence).
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
 AUTHORS Ishibashi, T., Kanehori, K., Yosida, M., Watanabe, S., Ishida, S.,
 Ono, Y., Horiuchi, T., Hiraoka, S., Murakawa, K., Takiguchi, S.,
 Kusano, J., Watanabe, M., Fujimori, K., Tanai, H., Ishida, M.,
 Yasushita, H., Chiba, Y., Sugiyama, T., Irie, R., Otsuki, T., Sato, H.,
 Wakamatsu, A., Ishii, S., Yamamoto, J. I., Isono, Y., Kawai-Hio, Y.,
 Saito, K., Nishikawa, T., Kimura, K., Matsuo, K., Nakamura, Y.,
 Sekine, M., Kikuchi, H., Kanda, K., Nagatsuma, M., Takahashi-Fujii, A.,
 Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S.,
 Nagahari, K., Masuno, Y., Nagai, K. and Isegai, T.
 NEDO human cDNA sequencing project
 Unpublished
 2 (bases 1 to 2684)
 Isegai, T., Otsuki, T. and Sugiyama, T.
 Direct Submission
 Submitted (24-OCT-2001) Takao Isegai, Helix Research Institute,
 Genomics Laboratory; 1532-3 Yana, Katsuzuu, Chiba 292-0812, Japan
 (E-mail: genomics@hri.co.jp, Tel: 81-438-52-9975, Fax: 81-438-52-9986)

COMMENT

NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5' - & 3' end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

FEATURES

Source

Location/Qualifiers

1. 2684
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PROST2005426"
/issue_type="prestate"
/clone_id="PROST2"
/note="Cloning vector: pME18SFL3"
56. 952
/note="unnamed protein product"
/protein_start=1
/protein_id="BAB71304.1"
/db_xref="GI:16552420"
/translation="MATTGHTPPFYAGKPTPTMTTASTIMFTLATPTVILPG
IRKSTRILFWLRLVSTLFTGATLTPVQLETTINNEFTWLGENVAYEYAALE
KGLPDPVLYIAEKFTPSPCGLRYRLAGHTSAMLWVAFCMLANVMSPVLY
GGYMLATGIFOLALFESMATSLTSPCLHIGASVLTTHGPAFWTLTGLLCVL
LGLAMAVAHMOPHRKAFPNOSVDEDPMLTLEMSPEBGILLSPRYSNADSPRSQDPL
SEASTVAYKEAHNPDPDCAI"

CDS

ORIGIN

Query Match

27.2%; Score 392; DB 9; Length 2684;

Best Local Similarity 100.0%; Pred. No. 1.2e-205;

Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1027 GGGCTCTGTCCTCTCTCGAGAGGCGGTGAGTCCAGTATGTCGCGCCAGC 1086
2243 GGGCTCTGTCCTCTCTCGAGAGGCGGTGAGTCCAGTATGTCGCGCCAGC 2284
1087 GCTCTTGCACCTCTTGCACCAAGCGCCAGAGTGCAGCCAGAGAGAGGCGCTCA 1146
2283 GCTCTTGCACCTCTTGCACCAAGCGCCAGAGTGCAGCCAGAGAGAGGCGCTCA 2224
1147 CCTCTTATCTCGGCGGACCACTGCAAGAGAGCGCTCTCCCACTTAAATGATAC 1206
2223 CCTCTTATCTCGGCGGACCACTGCAAGAGAGCGCTCTCCCACTTAAATGATAC 2164
1207 ACCACTAAGCTGTAGAGGAGGACCAATCTGACTCTTCCCGCCTTGGACATCGCAG 1266
2163 ACCACTAAGCTGTAGAGGAGGACCAATCTGACTCTTCCCGCCTTGGACATCGCAG 2104
1267 CCGGGAAGCAATGCGCCGCGCAGAGCTTGGGCGAGAGCTCCAGAAAGGACATGACCGCT 1326
2103 CCGGGAAGCAATGCGCCGCGCAGAGCTTGGGCGAGAGCTCCAGAAAGGACATGACCGCT 2044
1327 GCTGGCGGAGAGCTCGGACATCCGAGGAGACCAAGGAAAGTCTCTGGGCGCATCTGA 1386
2043 GCTGGCGGAGAGCTCGGACATCCGAGGAGACCAAGGAAAGTCTCTGGGCGCATCTGA 1984
1387 AATTAACCTTTTCTTTTCTTTTAAAAA 1418
1983 AATTAACCTTTTCTTTTCTTTTAAAAA 1952

RESULT 8

LOCUS

DEFINITION

AC009700 156534 bp DNA linear HTG 20-APR-2000
Homo sapiens chromosome 15 clone RP11-163p10 map 15, WORKING DRAFT

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

TITLE

JOURNAL

COMMENT

REFERENCE

1 (bases 1 to 156534)
Bairren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 15, clone RP11-163p10
Unpublished
2 (bases 1 to 156534)
Bairren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baker, J., Baldwin, J., Barna, N., Beckler, R., Benn, J., Brown, A.,
Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., Dearellano, K., Depayre, E., Devon, K., Dewar, K.,
Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C.,
Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G.,
Hague, B., Heath, A., Horton, L., Howland, J. C., Jones, C., Kamm, L.,
Karas, N., Lebecky, J., Lieu, C., Locke, K., MacDonald, P.,
Margis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J.,
Meldrum, J., Molla, M., Morris, W., Morrow, J., Myhalachy, J.,
Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,
Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,
Tessier, S., Torruella-Miller, J., Vassiliev, H., Vo, A., Wagner, A.,
Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.

Direct Submission
Submitted (28-AUG-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 20, 2000 this sequence version replaced gi:6056272.
All repeats were identified using RepeatMasker:
Solt, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIRB

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: 163.P.10

Center clone name: 163.P.10

Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960721

Consensus quality: 108837 bases at least Q40

Consensus quality: 130230 bases at least Q30

Consensus quality: 145227 bases at least Q20

Insert size: 15700; agarose-fp

Insert size: 15134; sum-of-contrigs

Quality coverage: 3.2 in Q20 bases; sum-of-contrigs

Quality coverage: 3.3 in Q20 bases; sum-of-contrigs

NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1169: contrig of 1169 bp in length
* 1170 1269: gap of 100 bp
* 1170 1269: gap of 100 bp
* 1270 3335: contrig of 2066 bp in length
* 1270 3335: gap of 100 bp
* 3336 3436: contrig of 2263 bp in length
* 3336 3436: gap of 100 bp
* 3436 5798: gap of 100 bp
* 5798 9216: contrig of 3418 bp in length
* 5798 9216: gap of 100 bp
* 9217 9317: gap of 100 bp
* 9317 15633: contrig of 6317 bp in length
* 9317 15633: gap of 100 bp
* 15634 15734: contrig of 6013 bp in length
* 15734 21747: gap of 100 bp
* 21747 21847: contrig of 8505 bp in length
* 21847 30352: gap of 100 bp
* 30352 30452: contrig of 6058 bp in length
* 30452 36509: gap of 100 bp
* 36509 36510: gap of 100 bp

```

* 36610 45379: contig of 8670 bp in length
* 45280 45379: gap of 100 bp
* 45380 56851: contig of 11472 bp in length
* 56852 67374: gap of 100 bp
* 67375 67474: gap of 100 bp
* 67475 79534: contig of 12060 bp in length
* 79535 95432: gap of 15798 bp in length
* 95433 95532: gap of 100 bp
* 95533 120801: contig of 25269 bp in length
* 120802 120901: gap of 100 bp
* 120902 156534: contig of 35633 bp in length.
Location/Qualifiers
1..156534
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone_lib="RP11-163P10"
/clone_1ib="RP11-163P10"
1..1169
/feature="assembly_fragment"
1270..3335
/feature="assembly_fragment"
3436..5698
/feature="assembly_fragment"
5799..9216
/feature="assembly_fragment"
9317..15633
/feature="assembly_fragment"
15734..21746
/feature="assembly_fragment"
21847..30351
/feature="assembly_fragment"
30452..36509
/feature="assembly_fragment"
clone_end:77
vector_side:left"
36610..45279
/feature="assembly_fragment"
45380..56851
/feature="assembly_fragment"
56952..67374
/feature="assembly_fragment"
67475..79534
/feature="assembly_fragment"
79635..95432
/feature="assembly_fragment"
95533..120801
/feature="assembly_fragment"
120902..156534
/feature="assembly_fragment"

```

```

FEATURES
source
1..156534
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone_lib="RP11-163P10"
/clone_1ib="RP11-163P10"
1..1169
/feature="assembly_fragment"
1270..3335
/feature="assembly_fragment"
3436..5698
/feature="assembly_fragment"
5799..9216
/feature="assembly_fragment"
9317..15633
/feature="assembly_fragment"
15734..21746
/feature="assembly_fragment"
21847..30351
/feature="assembly_fragment"
30452..36509
/feature="assembly_fragment"
clone_end:77
vector_side:left"
36610..45279
/feature="assembly_fragment"
45380..56851
/feature="assembly_fragment"
56952..67374
/feature="assembly_fragment"
67475..79534
/feature="assembly_fragment"
79635..95432
/feature="assembly_fragment"
95533..120801
/feature="assembly_fragment"
120902..156534
/feature="assembly_fragment"

```

ORIGIN

```

Query Match      27.2%; Score 392; DB 2; Length 156534;
Best Local Similarity 100.0%; Pred. No. 1.2e-205;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1027 GGGGCTGCTGCTCTTCTCTCGAGAGGCGCGGTGAGTCTCACTATCTTCGCCCGCCAGC 1086
Db 96131 GGGGCTGCTGCTCTTCTCTCGAGAGGCGCGGTGAGTCTCACTATCTTCGCCCGCCAGC 96190
QY 1087 GCTCTTGCACCTCTTGCACCAAGCGCAAGCACTGAGCGCAGAGAGAGAGGCGGCTCA 1146
Db 96191 GCTCTTGCACCTCTTGCACCAAGCGCAAGCACTGAGCGCAGAGAGAGAGGCGGCTCA 96250
QY 1147 CTTCTTATCTCTGCGGACCACTGCAAGCAGGCGGCTCTCCGAGACTTAATAAGTATC 1206
Db 96251 CTTCTTATCTCTGCGGACCACTGCAAGCAGGCGGCTCTCCGAGACTTAATAAGTATC 96310
QY 1207 ACCTAATCTGAGAGGAGGAGCAATCTGAGTCTCTCCGCGCTTGAGCATGCGAGG 1266

```

```

Db 96311 ACCACTAATCTGAGAGGAGGAGCAATCTGAGTCTCTCCGCGCTTGAGCATGCGAGG 96370
QY 1267 CCAGGAGCAGTGTCCCGCCAGGCGCTTGGGCGCAGAGAGCTTCAGAGAGGCACTGAGCGCT 1326
Db 96371 CCAGGAGCAGTGTCCCGCCAGGCGCTTGGGCGCAGAGAGCTTCAGAGAGGCACTGAGCGCT 96430
QY 1327 GCTGGCGGAGGCGCTCGACATCCGACAGGACACCGAGAAAGTCTCTGGGCGGATCTGTA 1386
Db 96431 GCTGGCGGAGGCGCTCGACATCCGACAGGACACCGAGAAAGTCTCTGGGCGGATCTGTA 96490
QY 1387 AATAACCTTTTCTTTTCTTTTCTTTTAAAAA 1418
Db 96491 AATAACCTTTTCTTTTCTTTTCTTTTAAAAA 96522

```

RESULT 9

AC087790

LOCUS

DEFINITION

AC087790.2

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

```

2 (bases 1 to 171444)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Boucknight,B., Brown,A.,
Cammarata,J., Campopiano,A., Choepey,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArchiello,K., Dewar,K., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,
Gardina,S., Ginde,S., Govette,M., Graham,L., Grand-Pierre,N.,
Hagob,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., Larocque,K., Lamazares,R., Landers,T.,
Lenczky,J., Levine,R., Liu,G., Maclean,C., Macdonald,E.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheeler,R., Meldrum,J., Menus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Rise,C., Rogoy,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schnupack,R., Seaman,S., Severy,P.,
Souarez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Vei,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-JAN-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 16, 2001 this sequence version replaced gi:12408501.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRB
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L12451
Center clone name: 276.X.9
----- Summary Statistics
Sequencing vector: Plasmid, N/A; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731

```

* NOTE: This is a "working draft" sequence. It currently consists of 13 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of 'N', but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

FEATURES

```
misc_feature /note="assembly_fragment
137827. .170053
/note="assembly_fragment
170154. .171444
/note="assembly_fragment
clone_end:r7
vector_side:right"
```

RESULT 10

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
AK009700	<i>Homo sapiens</i> chromosome 15 clone RP11-163P10 map 15, WORKING DRAFT	AC009700.4	GI:7622346	HTG; HTGS; PHASE1; HTGS_DRAFT.	Homo sapiens (human)	<i>Homo sapiens</i>
REFERENCE	Bukacynska, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	1 (bases 1 to 156534)				
AUTHORS	Birren, B., Linton, L., Nusbaum, C. and Lander, E.	2 (bases 1 to 156534)				
TITLE	Unpublished					
JOURNAL						
REFERENCE						
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Brown, A., Baker, J., Balwin, J., Barna, N., Beckerly, R., Benu, J., Brown, A., Castle, A., Cenny, J., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Depayre, B., Devon, K., Dewar, K., Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C., Funks, R., Gage, D., Galagan, J., Garrow, S., Gilbert, D., Grant, G., Hango, B., Heath, A., Horton, L., Howland, J. C., Jones, C., Kam, L., Karatas, A., Lehoczy, J., Lien, C., Locke, K., MacDonald, P., Maratsis, N., McEwan, P., McGuirk, A., McKernan, K., McLaughlin, J., Melidim, J., Mollia, M., Morris, W., Morrow, J., Mychaleckyj, J., Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B., Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P., Strange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A., Tesfaye, S., Tornella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,					

ACCESSION 257552 GI:1028783
 VERSION CPG island; genomic MseI fragment.
 KEYWORDS Homo sapiens (human)
 SOURCE
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1
 AUTHORS Cross, S.H., Charlton, J.A., Nan, X. and Bird, A.P.
 TITLE Purification of CpG islands using a methylated DNA binding column
 JOURNAL Nat. Genet. 6 (3), 236-244 (1994)
 MEDLINE 94282070
 PUBMED 8012384

REFERENCE 2 (bases 1 to 217)
 AUTHORS Dodsworth, S.J., Huckle, E., Wilkinson, P. and Micklem, G.
 TITLE Direct Substitution
 JOURNAL Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire, CB10 1RQ, England. E-mail contact: humquerry@sanger.ac.uk
 COMMENT Vector: pGEM-5Zf(-)
 Clones are available from the UK MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL: <http://www.hgmp.mrc.ac.uk/> for details or contact: biobelp@hgmp.mrc.ac.uk.
 Location/Qualifiers
 1..217
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="183h12"
 /sex="male"
 /tissue_type="blood"
 /clone_id="CGI-1"
 /dev_stage="adult"

ORIGIN
 Query Match 15.1%; Score 217; DB 9; Length 217;
 Best Local Similarity 100.0%; Pred. No. 1.4e-108; Indels 0; Gaps 0;
 Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1198 AATGATACACCACTAATCTGTAGAGGGAGCCCAATCTGACTCTTCCCGCCTTGGGA 1257
 Db 217 AATGATACACCACTAATCTGTAGAGGGAGCCCAATCTGACTCTTCCCGCCTTGGGA 158

QY 1258 CATGCGAGCGCGGAGACAGTCCCGCAGGCGCTGCGGACAGACTCTCAGAGAGGGA 1317
 Db 157 CATGCGAGCGCGGAGACAGTCCCGCAGGCGCTGCGGACAGACTCTCAGAGAGGGA 98

QY 1318 CTGAGCGCTGCTGCGCGGAGGCTCGGACATCCGACGAGGACGAGGAAAGTCTCTGAGG 1377
 Db 97 CTGAGCGCTGCTGCGCGGAGGCTCGGACATCCGACGAGGACGAGGAAAGTCTCTGAGG 38

QY 1378 CGATCTGTAAATAAAGCTTTTCTTTCTTTCTTTT 1414
 Db 37 CGATCTGTAAATAAAGCTTTTCTTTCTTTCTTTT 1

RESULT 12
 BC029819/c 1923 bp TRNA linear PRI 06-OCT-2003
 LOCUS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 DEFINITION Homo sapiens hypothetical protein FLJ32334, mRNA (cDNA clone MGC:35310 IMAGE:5177619), complete cds.
 ACCESSION BC029819
 VERSION BC029819.1 GI:20987583
 KEYWORDS MGC.
 SOURCE Homo sapiens (human)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 1923)
 AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.R., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schlier, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,

Diachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldi, M.P., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Robinson, J.S., Carninci, P., Prange, C., Raha, S., Loquiano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., Mobwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, J.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.W., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smalins, D.E., Scherch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL MEDLINE 22388257
 PUBMED 12477932

REFERENCE 2 (bases 1 to 1923)
 AUTHORS Strausberg, R.
 TITLE Direct Substitution
 JOURNAL Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
 DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louised, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Navevati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.llnl.gov>
 Series: IRK Plate: 51 Row: 0 Column: 3
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21389332.
 Location/Qualifiers
 1..1923
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:35310 IMAGE:5177619"
 /tissue_type="Brain, lung, Testis, adult, pooled whole"
 /clone_id="NIH_MGC_115"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 1..1923
 /gene="FLJ32334"
 /db_xref="locusid:90527"
 401..1852
 /codon_start=1
 /product="FLJ32334 protein"
 /protein_id="AA29819.1"
 /db_xref="GI:20987584"
 /db_xref="locusid:90527"
 /translation="MATLGTTPFVAGPKPTPMDTTLASITIMFLTALAPFVILPG
 IRKTLFPLILRYVTSITFGAIIIVNSESSEKSGVSTNSYKAPSESSADIGLO
 VGGVAVITITIGPVQINLETININDEFTWIRGENTAYRIALALEKGPDPVLYIAEK
 FTRSPGILYRQRLAGHTSAMLVAFICWLLAVVMSPVLTGGLGGLATGIFOL
 LALIFSMATSLSPGCHLIGASLVATHGPAFTVLTGLCVLLGLAMVAAHMQP
 HRKAFNDSVDPMLTWSPEGGILSPRYSDSPSQDIPLSBASSTKAVYRFP
 RLIVPDAVGLALPALSLPGLLAQARALLPGIRCPKAKESRLPSPVPMFGR
 EGGERMAEHTGSPRLRGRTGRIMRWGSKERRACGVFAVLPVLVNSGSKRPSCL

ORIGIN DLPKMDYRDRARAFPHLEPTPCVTSRHTPLI

Query Match 11.5%; Score 166; DB 9; Length 1923;
Best Local Similarity 100.0%; Pred. No. 2,7e-80;
Matches 166; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1216 CTGTAGAGGGGAGCCATCTGAGCTCTCCCGCCCTTGGGACATCGACGGCCGGGAAGC 1275
DB 1557 CTGTAGAGGGGAGCCATCTGAGCTCTCCCGCCCTTGGGACATCGACGGCCGGGAAGC 1498
QY 1216 AGTGCCTCCGAGGCTTGGGCGAGGAGCTCCAGGAAAGGCTCTGGGGCGAT 1381
DB 1497 AGTGCCTCCGAGGCTTGGGCGAGGAGCTCCAGGAAAGGCTCTGGGGCGAT 1438
QY 1336 AGGCTCCGAGGCTTGGGCGAGGAGCTCCAGGAAAGGCTCTGGGGCGAT 1392
DB 1437 AGGCTCCGAGGCTTGGGCGAGGAGCTCCAGGAAAGGCTCTGGGGCGAT 1392

RESULT 13 H5183A12R 218 bp DNA linear PRI 19-OCT-1995
LOCUS H. sapiens Cpg island DNA genomic MseI fragment, clone 183a12,
DEFINITION reverse read cpg183a12.rta.
ACCESSION Z59954.1 GI:1031867
VERSION Cpg island; genomic MseI fragment.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1
AUTHORS Cross, S.H., Charlton, J.A., Nan, X. and Bird, A.P.
TITLE Purification of Cpg islands using a methylated DNA binding column
JOURNAL Nat. Genet. 6 (3), 236-244 (1994)
MEDLINE 94282070
PUBMED 8012384
REFERENCE 2 (bases 1 to 218)
AUTHORS Dodsworth, S.J., Huckle, E., Wilkinson, P. and Micklem, G.
TITLE Direct Submission
SUBMITTED (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, England. E-mail contact: humquery@sanger.ac.uk
COMMENT Vector: pGEM-5ZF(-)
Clones are available from the UK MRC Human Genome Mapping Project
Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:
http://www.hgmp.mrc.ac.uk/ For details
or contact: biohelp@hgmp.mrc.ac.uk.
Location/Qualifiers

FEATURES
source
1. .218
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="183a12"
/sex="male"
/tissue_type="blood"
/clone_lib="CGI-1"
/dev_stage="adult"

ORIGIN
Query Match 9.4%; Score 136; DB 9; Length 218;
Best Local Similarity 100.0%; Pred. No. 1.2e-63;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1268 CGGAGACAGTGCCTCCGAGGCTTGGGCGAGGAGCTCCAGGAAAGGCTGAGCGCTG 1327
DB 72 CGGAGACAGTGCCTCCGAGGCTTGGGCGAGGAGCTCCAGGAAAGGCTGAGCGCTG 131
QY 1328 CTGGCGGAGGCTTGGGAGCTCCGAGGAGCTCCAGGAAAGTCTCTCTGGGGCGATCTGTAA 1387
DB 132 CTGGCGGAGGCTTGGGAGCTCCGAGGAGCTCCAGGAAAGTCTCTCTGGGGCGATCTGTAA 191
QY 1388 ATAAACCTTTTCT 1403

DB 192 ATAAACCTTTTCT 207

RESULT 14 BC004161 1725 bp mRNA linear PRI 12-NOV-2003
LOCUS Homo sapiens transmembrane 4 superfamily member tetraaplan NET-7,
DEFINITION mRNA (CDNA clone MGC:2447 IMAGE:2958221), complete cds.
ACCESSION BC004161
VERSION BC004161.2 GI:33872773
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1725)
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Altschul, S.F., Zeeberg, B., Bietow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Wax, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Donaldson, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.U., Malek, U.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Faney, J., Helton, E., Ketterman, M., Madan, A., Rodrigues, S.,
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywicki, M.I., Skalsky, J., Smalls, D.E.,
Scherer, A., Schein, J.E., Jones, S.T. and Marra, M.A.
TITLE Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE 22388257
PUBMED 12477932
REFERENCE 2 (bases 1 to 1725)
AUTHORS Strausberg, R.
TITLE Direct Submission
SUBMITTED (01-MAR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK NIH-MGC Project URL: http://mgc.ncl.nih.gov
On Aug 19, 2003 this sequence version replaced gi:13278776.
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: DCTD/DTF
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Ketterman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whitting

FEATURES
source
1. .1725
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/db_xref="IMAGE:2958221"
/tissue="MGC:2447 IMAGE:2958221"
/tissue_type="Kidney, renal cell adenocarcinoma"
/clone_lib="NIH_MGC_14"

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRL Plate: 2 Row: K Column: 4
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 21264576.
Location/Qualifiers

```

gene          /lab_host="DH10B-R"
              /note="vector: POTB7"
              1..1725
              /gene="NEF-7"
              /db_xref="LOCUSID:23555"
              120..1004
              /codon_start=1
              /product="tetraspan NEF-7"
              /protein_id="AA04161.1"
              /db_xref="GI:13278777"
              /db_xref="LOCUSID:23555"
              /translation="MPRGSEQVRYCARPSYLMKFSLIYTVFWLIGALVLSYGIY
              AVEYRKXTLSAFLAPAIILILGVWFMVSPFGLASLDNLYLQAFMYILGIC
              LIMELIGVVALFRNOTDFLNDNRKREIENYDDLPKINMDVCKRFGCCGEDY
              RDKSKQYHDCSAPGACGVPYTCIRNTETVNMVCGYKTDIKERSVQDVIYVRG
              CTNAVITIMENDNTIMAGITLILPQFLGVLLTLITRVEDIIMERSVTDGLLPG
              AKSVEAAGTGCCLCYPN"
              174..854
              /note="transmembrane4; Region: Tetraspanin family"
              /db_xref="CDD:pfam00335"

misc_feature
174..854
/note="transmembrane4; Region: Tetraspanin family"

ORIGIN
Query Match      2.7%; Score 39; DB 9; Length 1725;
Best Local Similarity 100.0%; Pred. No. 7.8e-10;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1403 TTTGTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1441
Db      1683 TTTGTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1721

RESULT 15
LOCUS      BC003157      1726 bp      mRNA      linear      PRI 03-OCT-2003
DEFINITION Homo sapiens transmembrane 4 superfamily member tetraspan NEF-7,
              mRNA (cDNA clone MGC:4120 IMAGE:2958221), complete cds.
ACCESSION   BC003157
VERSION     BC003157.1 GI:13111968
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1726)
Strausberg,R.L., Fengold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Ditchenko,L., Marusik,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Cassavant,T.L.,
Schneer,T.E., Brownstein,M.J., Uebin,T.B., Teshnyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Muliahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gutaralme,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettman,M., Madan,A.C., Shavchenko,Y.,
Sanche,A., Whiting,M., Madan,A., Young,A.C., Shavchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,C.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smallus,D.E.,
Scherer,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
2388257
JOURNAL
MEDLINE
PUBMED      2 (bases 1 to 1726)
12477932
REFERENCE
AUTHORS      Strausberg,R.
TITLE        Direct Submission
JOURNAL      Submitted (13-FEB-2001) National Institutes of Health, Mammalian
              Gene Collection (MGC), Cancer Genomics Office, National Cancer
              Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

```

```

REMARK
COMMENT
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: DCTD/DTF
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@gscc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susana Chan, Readman Chiu, Chris Fiehl, Erin Gariand, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Parbhu, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smallus, Michael Smith, Loraine Spence, Jeff Stott,
Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov
Series: IRL Plate: 9 Row: n Column: 20
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 21264576.
Location/Qualifiers
1..1726
/organism="Homo sapiens"
/mol_type="RNA"
/db_xref="taxon:9606"
/clone="MGC:4120 IMAGE:2958221"
/tissue_type="Kidney, renal cell adenocarcinoma"
/clone_id="NIH-MGC-14"
/lab_host="DH10B-R"
/note="vector: POTB7"
1..1726
/gene="NEF-7"
/db_xref="LOCUSID:23555"
120..1004
/codon_start=1
/product="tetraspan NEF-7"
/protein_id="AA03157.1"
/db_xref="GI:13111969"
/db_xref="LOCUSID:23555"
/translation="MPRGSEQVRYCARPSYLMKFSLIYTVFWLIGALVLSYGIY
AVEYRKXTLSAFLAPAIILILGVWFMVSPFGLASLDNLYLQAFMYILGIC
LIMELIGVVALFRNOTDFLNDNRKREIENYDDLPKINMDVCKRFGCCGEDY
RDKSKQYHDCSAPGACGVPYTCIRNTETVNMVCGYKTDIKERSVQDVIYVRG
CTNAVITIMENDNTIMAGITLILPQFLGVLLTLITRVEDIIMERSVTDGLLPG
AKSVEAAGTGCCLCYPN"
180..896
/note="transmembrane4; Region: Tetraspanin family"
/db_xref="CDD:pfam00335"

misc_feature
180..896
/note="transmembrane4; Region: Tetraspanin family"
/db_xref="CDD:pfam00335"

ORIGIN
Query Match      2.7%; Score 39; DB 9; Length 1726;
Best Local Similarity 100.0%; Pred. No. 7.8e-10;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1403 TTTGTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1441
Db      1683 TTTGTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1721

Search completed: February 24, 2004, 01:57:04
Job time : 5689.97 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 19:50:14 / Search time 582.643 Seconds
10506.704 Million cell updates/sec

Title: US-09-936-456-1
Perfect score: 1441

Sequence: 1 aaagtaacgctcacagacag.....aaaaaaaaaaaaa 1441

Scoring table: OLIGO_NUC
Gapop 60.0, Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size: 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: N_Geneseq_29Jan04:*

1: geneseqn1808:*\n2: geneseqn1908:*\n3: geneseqn2000s:*\n4: geneseqn2001as:*\n5: geneseqn2001bs:*\n6: geneseqn2002s:*\n7: geneseqn2003as:*\n8: geneseqn2003bs:*\n9: geneseqn2003cs:*\n10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1441	100.0	1441	3	AA94623 Human CAS
2	1413	98.1	1421	3	AA96505 Human CAS
3	1336	92.7	1451	6	ABD19228 Human CDN
4	1285	89.2	1474	6	ABL90535 Human pol
5	996	69.1	1460	3	AA95564 Human sec
6	996	69.1	1460	7	AB267235 Human sec
7	996	69.1	1460	7	AB273640 Secreted
8	996	69.1	1460	9	ADC20289 Human sec
9	425	29.5	1498	3	AA94624 Human CAS
10	405	28.1	5033	7	ABD19167 Human CDN
11	405	28.1	5033	7	AB268115 Human sec
12	405	28.1	5033	7	AB274587 Secreted
13	405	28.1	5033	9	ADC21005 Human sec
14	392	27.2	2684	7	AD453627 Human cod
15	392	27.2	406	5	AA56563 DNA encod
16	59	4.1	512	4	AA36165 Human col
17	38	2.6	427	4	AA187437 Human pol
18	38	2.6	427	4	AA187437 Human pol
19	38	2.6	2165	3	AA90468 Human unc
20	38	2.6	5642	6	ABL32869 Human imm
21	38	2.6	10480	6	AA134201 Human imm
22	36	2.5	896	4	AAH33168 Human col
23	36	2.5	2255	2	AA964698 Human IFN

C	24	35	2.4	297	5	ABV61920 Human pro
C	25	35	2.4	364	4	AA193544 Human pol
C	26	35	2.4	436	5	ABV35678 Human pro
C	27	35	2.4	437	5	ABV44483 Human pro
C	28	35	2.4	450	6	AB208656 Human leu
C	29	35	2.4	453	4	AA187364 Human pol
C	30	35	2.4	462	5	ABV34102 Human pro
C	31	35	2.4	462	5	ABV42965 Human pro
C	32	35	2.4	464	6	AB208188 Human leu
C	33	35	2.4	504	5	ABV56694 Human pro
C	34	35	2.4	572	5	ABV61258 Human pro
C	35	35	2.4	629	5	ABV12980 Human pro
C	36	35	2.4	644	4	AA123651 Human bre
C	37	35	2.4	748	4	AAH71460 Human cer
C	38	35	2.4	830	5	ABV14600 Human pro
C	39	35	2.4	1034	3	AA252527 Human sec
C	40	35	2.4	1639	5	AA252527 Nucleotid
C	41	35	2.4	2857	9	AD445624 Human gen
C	42	35	2.4	2857	9	AD60246 Human gen
C	43	35	2.4	2857	9	AD60248 Human gen
C	44	35	2.4	2857	9	AD60252 Human gen
C	45	35	2.4	2857	9	AD60250 Human gen

ALIGNMENTS

RESULT 1	AAA94623	standard; DNA; 1441 BP.
ID	AAA94623	
XX	AAA94623	
XX	11-JAN-2001	(first entry)
DT	XX	
XX	Human CASB618 coding sequence.	
DE	XX	
XX	Human CASB618: chromosome 15; tumour; vaccine; cancer; ovarian;	
KW	colony; autoimmune disease; HLA_A0201; ss.	
XX	XX	
XX	Homo sapiens.	
OS	XX	
XX	Key	Location/Qualifiers
FT	CD5	259..1221
FT	FT	/tag=a
FT	FT	/product="Human CASB618"
XX	XX	
XX	WO200053748-A2.	
XX	PD	14-SEP-2000.
XX	XX	
PF	09-MAR-2000; 2000WO-EP002048.	
XX	XX	
PR	11-MAR-1999; 99GB-00005607.	
PR	01-SEP-1999; 99GB-00020590.	
XX	XX	
PA	(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.	
XX	XX	
PI	Bruck CEM, Caesart J, Coche T, Vinals Y De Bassolec;	
XX	XX	
DR	WPI, 2000-572268/53.	
DR	P-PSDB; AAB63325.	
XX	XX	
PT	New human CASB618 polypeptide, useful as a vaccine for prophylactic and	
PT	therapeutic treatment of cancers, particularly ovarian or colon cancer,	
PT	autoimmune diseases and related conditions.	
XX	XX	
PS	Claim 13, Page 61, 76pp, English.	
XX	XX	
CC	The present sequence is the coding sequence of human CASB618 protein. The	
CC	gene for human CASB618 is thought to be located on chromosome 15. The	
CC	protein encoded by the present sequence and epitopes of the CASB618	
CC	protein (see AAB26327 to AAB26399) are useful in diagnosing the occurrence	

CC of tumour cells and in vaccines for prophylactic and therapeutic
CC treatment of cancers, particularly ovarian or colon cancer, autoimmune
CC diseases and related conditions

Sequence 1441 BP; 289 A; 466 C; 392 G; 294 T; 0 U; 0 Other;

Query Match	100.0%;	Score 1441;	DB 3;	Length 1441;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1441; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY	1	AAAGTAAAGCGCTCAAGACAGTGAAGAAATGTTGCTCGCGCGGCTGAAGAAAATCTGTGCG	60
Db	1	AAAGTAAAGCGCTCAAGACAGTGAAGAAATGTTGCTCGCGCGGCTGAAGAAAATCTGTGCG	60
QY	61	GTACCAACCCCAAGAGCTTTAGAGAGCAACCCACTCCACGCTTCTCTTACGAGAGGTGCA	120
Db	61	GTACCAACCCCAAGAGCTTTAGAGAGCAACCCACTCCACGCTTCTCTTACGAGAGGTGCA	120
QY	121	GGAGCTAAGACTTCAACAGGCCCACTCGGCTCCAGGCTGTGTAAGCAAGAGCGCAAGGAC	180
Db	121	GGAGCTAAGACTTCAACAGGCCCACTCGGCTCCAGGCTGTGTAAGCAAGAGCGCAAGGAC	180
QY	181	GGCGCTCTCCCGGTCCAGGCAAGCCCACTTGCTGAGCTTGCTGACCCGCTGGGTCAAC	240
Db	181	GGCGCTCTCCCGGTCCAGGCAAGCCCACTTGCTGAGCTTGCTGACCCGCTGGGTCAAC	240
QY	241	ACTCGGCGGGCGTGGACAGCATACCTGTGGAAACGGGCTACTGCTTTTAAACCCCAAGCC	300
Db	241	ACTCGGCGGGCGTGGACAGCATACCTGTGGAAACGGGCTACTGCTTTTAAACCCCAAGCC	300
QY	301	CGGCAATGCCGAGAGCTTACAGCTTCACTGCTCATCTGTATTCTAGTGTTTTGGCTCTA	360
Db	301	CGGCAATGCCGAGAGCTTACAGCTTCACTGCTCATCTGTATTCTAGTGTTTTGGCTCTA	360
QY	361	GCAGCAAGCTTCTGCTCTCATCTTGGCGGGGATCCGTGGCACTCGCGCTGGTTTGGTTG	420
Db	361	GCAGCAAGCTTCTGCTCTCATCTTGGCGGGGATCCGTGGCACTCGCGCTGGTTTGGTTG	420
QY	421	GTAGAGAGTCTTCTCAGTCTGTATCAATPAGCGGAGAAATGGGCTGGTGCACCTCAGTGA	480
Db	421	GTAGAGAGTCTTCTCAGTCTGTATCAATPAGCGGAGAAATGGGCTGGTGCACCTCAGTGA	480
QY	481	GAATGATTCGTGGGTACAGTGAACCAACCATCTCTCAAAAGGCTTCAACGCAAGCGCGC	540
Db	481	GAATGATTCGTGGGTACAGTGAACCAACCATCTCTCAAAAGGCTTCAACGCAAGCGCGC	540
QY	541	GTTACACGCGCGTGGTGGTGTGCTGGTGGGCGCTGAGAGGCAATTATTAACATCAAGGG	600
Db	541	GTTACACGCGCGTGGTGGTGTGCTGGTGGGCGCTGAGAGGCAATTATTAACATCAAGGG	600
QY	601	ACCCCAAGTCATCAGCTGAACAGAACCAATTGATCAACAAGACAGTTCACCTGGCGTCTG	660
Db	601	ACCCCAAGTCATCAGCTGAACAGAACCAATTGATCAACAAGACAGTTCACCTGGCGTCTG	660
QY	661	AAAGAGAAATTAACGCGCGCGAGTAAACCGCAACGACCTGGAGAAAGGGGCTCGCGAACCAAGT	720
Db	661	AAAGAGAAATTAACGCGCGCGAGTAAACCGCAACGACCTGGAGAAAGGGGCTCGCGAACCAAGT	720
QY	721	CTGTACCTGGCGAGAAATTCACACCGAGTAGCCCTTTCGGGCTGTACCAACAGTACAC	780
Db	721	CTGTACCTGGCGAGAAATTCACACCGAGTAGCCCTTTCGGGCTGTACCAACAGTACAC	780
QY	781	CTGGCGGGAGACTTAAGCCTTCGGGCAACGCTAAGGATGGGCTGTCTGTCTGTGGCTCTCTCC	840
Db	781	CTGGCGGGAGACTTAAGCCTTCGGGCAACGCTAAGGATGGGCTGTCTGTCTGTGGCTCTCTCC	840
QY	841	AAAGTGTGCTTTCACAGCGCGGCGCGCTTACAGGAGCGCTGGACATGCTGAACACAGGA	900
Db	841	AAAGTGTGCTTTCACAGCGCGGCGCGCTTACAGGAGCGCTGGACATGCTGAACACAGGA	900
QY	901	GCCTTGGCGCTCTTGGGGATCTTCCGCTTGGCTCTCATCTTAAGCTGCGGCTCTGCTCGG	960
Db	901	GCCTTGGCGCTCTTGGGGATCTTCCGCTTGGCTCTCATCTTAAGCTGCGGCTCTGCTCGG	960

QY	161	CTCCGCGTAAAGGCTCTCCGAGCTACACACTAGATAGAGGCGCGCTTCCTGGGTCACGGTC	1020
Db	961	CTCCGCGTAAAGGCTCTCCGAGCTACACACTAGATAGAGGCGCGCTTCCTGGGTCACGGTC	1020
QY	1021	GCACCGGCGTCTCTGGCTTCTTCCTCGAGGAGGAGCCGTGATGTCCTCCAGTATGTTCCG	1080
Db	1021	GCACCGGCGTCTCTGGCTTCTTCCTCGAGGAGGAGCCGTGATGTCCTCCAGTATGTTCCG	1080
QY	1081	CCGACGGCTTTCGGACCTTCCTGGACCAAGGCGCAGAGATGCGACGACCGAGAGAGG	1140
Db	1081	CCGACGGCTTTCGGACCTTCCTGGACCAAGGCGCAGAGATGCGACGACCGAGAGAGG	1140
QY	1141	GCGTCACTCTTATCTCTGGGCAACCACCTGACACAGACAGCGCTTCTCCAGACTTAA	1200
Db	1141	GCGTCACTCTTATCTCTGGGCAACCACCTGACACAGACAGCGCTTCTCCAGACTTAA	1200
QY	1201	TGTATTCACCACTAAACCTGGAGAGGAGGACCCAACTGTGATCTCTTCCCGGCTTGGAGCAT	1260
Db	1201	TGTATTCACCACTAACTGTGAGAGGAGGACCCAACTGTGATCTCTTCCCGGCTTGGAGCAT	1260
QY	1261	CGCAGGCGCGGAGAGCAGTGCCTCGCAGGCTCGGACAGAGAGCTTCAGAGAGGCACTG	1320
Db	1261	CGCAGGCGCGGAGAGCAGTGCCTCGCAGGCTCGGACAGAGAGCTTCAGAGAGGCACTG	1320
QY	1321	AAGCGTGTCTGGGCGAGGCTTCGGACATCCGACAGGACCAAGAGGAAAGTCTCTCGGGGCGA	1380
Db	1321	AAGCGTGTCTGGGCGAGGCTTCGGACATCCGACAGGACCAAGAGGAAAGTCTCTCGGGGCGA	1380
QY	1381	TCTGTAAATAAACCTTTTTCCTTTGTTTTAAAAA	1440
Db	1381	TCTGTAAATAAACCTTTTTCCTTTGTTTTAAAAA	1440
QY	1441	A 1441	
Db	1441	A 1441	

	RESULT 2
XX	AAA96505
ID	AAA96505 standard; cDNA, 1421 BP.
XX	
AC	AAA96505;
XX	
DT	08-FEB-2001 (first entry)
XX	
DE	cDNA encoding a human transmembrane protein.
XX	
KW	Human; transmembrane protein; cell proliferation disorder; myeloma;
KW	reproductive disorder; smooth muscle disorder; neurological disorder;
KW	arteriosclerosis; leukæmia; acquired immunodeficiency syndrome; AIDS;
KW	allergy; ovulatory defect; angina; hypertension; stroke; epilepsy;
KW	Alzheimer's disease; Tourette's disorder; ss.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	Location/Qualifiers
FT	CDS
FT	267..1229
XX	/tag= a
XX	
PN	WO200056891-A2.
XX	
XX	28-SEP-2000.
PD	
XX	
PF	22-MAR-2000; 2000MO-US007817.
XX	
PR	22-MAR-1999; 99US-012537P.
PR	16-JUN-1999; 99US-013956SP.
XX	
PA	(INCY-) INCYTE PHARM INC.
XX	
PI	Xue H, Lal P, Tang YT, Hillman JL, Reddy R, Bardman O,
	Baughn KR, Lu DM, Azimzai Y, Yang J;

XX WPI, 2000-579485/54.
 DR P-PSDB; AAB18992.
 XX New human transmembrane proteins are used to treat a disease or condition
 PT associated with decreased expression of functional HTPM e.g. Tourette's
 disorder, angina and leukemia.
 PS Claim 4, Page 129, 130pp; English.
 XX The present sequence encodes a human transmembrane proteins (HTMP).
 CC Agonists and antagonists of the protein are used to treat a disease or
 CC condition associated with overexpression of the protein. Diseases and
 CC conditions which can be treated include cell proliferative,
 CC immunological, reproductive, smooth muscle and neurological disorders
 CC e.g. arteriosclerosis, myeloma, leukemia, acquired immunodeficiency
 CC syndrome (AIDS), allergies, ovulatory defects, angina, hypertension,
 CC stroke, Alzheimer's disease, epilepsy and Tourette's disorder. The
 CC polynucleotides may be used to detect and quantify gene expression in
 CC biopsied tissues where protein expression may be correlated with disease
 CC e.g. to determine absence, presence or excess expression of HTPM or to
 CC monitor regulation of HTPM expression during therapeutic intervention
 XX
 SQ Sequence 1421 BP, 264 A; 468 C; 395 G; 294 T; 0 U; 0 Other;
 Query Match 98.1%; Score 1413; DB 3; Length 1421;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 661 AAAGAAATTACGCGCGGAGTACGGAACGCACTGGAAGGGGCTGCGGACCCAGT 720
 DB 669 AAAGAAATTACGCGCGGAGTACGGAACGCACTGGAAGGGGCTGCGGACCCAGT 728
 QY 721 CTCTACCTGGCGGAGAGTTTCAACCGAGTACCTTGGCGGCTGTACACAGTACAC 780
 DB 729 CTCTACCTGGCGGAGAGTTTCAACCGAGTACCTTGGCGGCTGTACACAGTACAC 788
 QY 781 CTGGCGGAGACATACGCGGCTGCGGAGTACGGAACGCACTGGAAGGGGCTGCGG 840
 DB 789 CTGGCGGAGACATACGCGGCTGCGGAGTACGGAACGCACTGGAAGGGGCTGCGG 848
 QY 841 AACGTGCTCTTCACAGCGCGGCGGCTCTTACGAGAGCTTGGCACTGTACACAGG 900
 DB 849 AACGTGCTCTTCACAGCGCGGCGGCTCTTACGAGAGCTTGGCACTGTACACAGG 908
 QY 901 GCGTGGCGCTCTTGGGAGTCTTGGCTTGGGCTTCCATCTAGAGTGGCGGCTGCGG 960
 DB 909 GCGTGGCGCTCTTGGGAGTCTTGGCTTGGGCTTCCATCTAGAGTGGCGGCTGCGG 968
 QY 961 CTGGCGTACGCTCTTGGGAGTCTTGGCTTGGGCTTCCATCTAGAGTGGCGGCTGCGG 1020
 DB 969 CTGGCGTACGCTCTTGGGAGTCTTGGCTTGGGCTTCCATCTAGAGTGGCGGCTGCGG 1028
 QY 1021 GCAACGCGGCTCTTGGGAGTCTTGGCTTGGGCTTCCATCTAGAGTGGCGGCTGCGG 1080
 DB 1029 GCAACGCGGCTCTTGGGAGTCTTGGCTTGGGCTTCCATCTAGAGTGGCGGCTGCGG 1088
 QY 1081 CCCAGGCGCTCTTGGGAGTCTTGGCTTGGGCTTCCATCTAGAGTGGCGGCTGCGG 1140
 DB 1089 CCCAGGCGCTCTTGGGAGTCTTGGCTTGGGCTTCCATCTAGAGTGGCGGCTGCGG 1148
 QY 1141 GCGTACCTTCTTACCTGCGGAGCACTGACACAGAGCGGCTTCCATCTAGAGTGG 1200
 DB 1149 GCGTACCTTCTTACCTGCGGAGCACTGACACAGAGCGGCTTCCATCTAGAGTGG 1208
 QY 1201 TGTATCAACATACCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1260
 DB 1209 TGTATCAACATACCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1268
 QY 1261 CGCAGGCGGAGAGCACTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1320
 DB 1269 CGCAGGCGGAGAGCACTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1328
 QY 1321 AGCGCTGCTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1380
 DB 1329 AGCGCTGCTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1388
 QY 1381 TCTGTAAATAAACCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1413
 DB 1389 TCTGTAAATAAACCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1421

RESULT 3
 ADD19228
 ID ADD19228 standard; cDNA; 1491 BP.
 XX
 AC ADD19228;
 XX
 DE Human cDNA from secreted protein gene 45.
 XX
 KW human secreted protein; cytosolic; antibacterial; virucide;
 KW neuroprotective; gynaecological; gastrointestinal; muscular; giant;
 KW cardiovascular; nephrotropic; antiinflammatory; muscular; giant;
 KW respiratory; gen; immunosuppressive; cerebroprotective; vasodilator;
 KW neotropic; antiallergic; cancer; bacterial infection; viral infection;
 KW neural disorder; immune system disorder; blood disorder;
 KW muscular disorder; reproductive disorder; gastrointestinal disorder;
 KW pulmonary disorder; cardiovascular disorder; renal disorder;
 KW inflammatory disorder; proliferative disorder; human; ss; gene.

OS Homo sapiens.
 XX WO2003052377-A2.
 XX
 XX 26-JUN-2003.
 XX
 XX 06-NOV-2002; 2002WO-US035606.
 XX
 XX 07-NOV-2001; 2001US-0331046P.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Ruben SM;
 XX WPI; 2003-533050/50.
 XX P-PSDB; ADD19303.
 XX
 XX New isolated nucleic acids encoding signal transduction pathway component
 PT polypeptides, useful for diagnosing, treating, and/or preventing
 PT disorders, such as cancer, infections, cardiovascular and inflammatory
 PT diseases.
 PT
 PS Claim 1; SEQ ID NO 55; 554bp; English.

XX The invention relates to an isolated nucleic acid molecule (cDNA)
 CC encoding a human secreted protein, representing one of 85 novel genes.
 CC Also included are recombinant vectors, host cells (expressing the
 CC protein), the secreted proteins (including their fragments, epitopes and
 CC homologues), an isolated antibody that binds specifically to the protein,
 CC diagnosing a pathological condition or susceptibility to a pathological
 CC condition (comprising determining the presence or absence of a mutation
 CC in the nucleic acid and diagnosing a condition based on the presence or
 CC absence of the mutation), diagnosing a pathological condition or
 CC susceptibility to a pathological condition (comprising determining the
 CC presence or amount of expression of the protein in a biological sample
 CC and diagnosing a condition based on the presence or amount of expression
 CC of the protein), preventing, treating or ameliorating a medical condition
 CC by administering the nucleic acid or protein to a mammalian subject,
 CC identifying a binding partner to the protein, the gene corresponding to
 CC the cDNA sequence, and identifying an activity in a biological assay
 CC (comprising expressing the nucleic acid in a cell, isolating the
 CC supernatant, detecting an activity in a biological assay and identifying
 CC the protein in the supernatant having the activity). The nucleic acids
 CC and proteins display the following activities: cytostatic, antibacterial,
 CC virucide, neuroprotective, gynecological, gastrointestinal, Gen,
 CC Cardiac, Cardiovascular-Gen, Nephrotropic, Antiinflammatory, Muscular-
 CC Gen, Respiratory-Gen, Immunosuppressive, Cerebroprotective, Vasectopic,
 CC Nootropic, Antiallergic. The methods and compositions of the present
 CC invention are useful for diagnosing, treating, preventing and/or
 CC prognosticating disorders related to the novel polypeptides, such as
 CC cancer, bacterial or viral infections, and neural, immune system, blood,
 CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
 CC renal, inflammatory or proliferative disorders (many examples of these
 CC diseases and disorders are given in the specification). The present
 CC sequence encodes a novel secreted protein of the invention.

XX Sequence 1491 BP; 293 A; 485 C; 410 G; 302 T; 0 U; 1 Other;

XX Query Match 92.7%; Score 1336; DB 9; Length 1491;

XX Best Local Similarity 99.9%; Pred. No. 0; Mismatches 2; Indels 0; Gaps 0;

XX Matches 1436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

XX 1 AAAGTAACGGCTACAGACAGTGAAGAAATAGTTTCGCTCGCGGCTAGAAAACTGTGCG 60
 Db 44 AAAGTAACGGCTACAGACAGTGAAGAAATAGTTTCGCTCGCGGCTAGAAAACTGTGCG 103
 QY 61 GTACCAACCCGAGGCGTTGAGAGACGCCCACTCCACGCTTCTTAAAGAGAGGTGA 120
 Db 104 GTACCAACCCGAGGCGTTGAGAGACGCCCACTCCACGCTTCTTAAAGAGAGGTGA 163
 QY 121 GGAATCAAGACTTACAGACAGCCCACTCGGTCCAGCCTTGTATAGCAAGAGAGCCAGAGAC 180
 Db 164 GGAATCAAGACTTACAGACAGCCCACTCGGTCCAGCCTTGTATAGCAAGAGAGCCAGAGAC 223

QY 181 GGGCTCTCCGCGCGTCCAGAGAGCCCGAGCTTGTGCTTGTCTGCGCGCGCTGGTGCAGC 240
 Db 224 GGGCTCTCCGCGCGTCCAGAGAGCCCGAGCTTGTGCTTGTCTGCGCGCGCTGGTGCAGC 283
 QY 241 ACTCGGCGCGGCTGACAGATGACCCTGTGGAACGGGGTACTGCTTTTAAACCCAGGCC 300
 Db 284 ACTCGGCGCGGCTGACAGATGACCCTGTGGAACGGGGTACTGCTTTTAAACCCAGGCC 343
 QY 301 CGGATGCGCGAGGCTTCAAGGTTCCATGCTCATGCTTATCTAGTGTTTTGTGCTTA 360
 Db 344 CGGATGCGCGAGGCTTCAAGGTTCCATGCTCATGCTTATCTAGTGTTTTGTGCTTA 403
 QY 361 GCAGCAACCTTCTGCTCATCTTTCGCGGGATCCGTGGCCACTCGCGCTGTGTTGTTG 420
 Db 404 GCAGCAACCTTCTGCTCATCTTTCGCGGGATCCGTGGCCACTCGCGCTGTGTTGTTG 463
 QY 421 GTGAGAGTCTTCTCAGTCTGTTTATAGGGGCAAAATGTGCTGTGCACTTACATGCA 480
 Db 464 GTGAGAGTCTTCTCAGTCTGTTTATAGGGGCAAAATGTGCTGTGCACTTACATGCA 523
 QY 481 GAATGTTCTGCGGTACAGTGAACCAACCAACATCTTACAAAGCTTTCAGCGAGCGCGC 540
 Db 524 GAATGTTCTGCGGTACAGTGAACCAACCAACATCTTACAAAGCTTTCAGCGAGCGCGC 583
 QY 541 GTTACAGCCCGTGTGCTGCTCTGTGCGCTGAGAGGCACTTAATTAATTAACACAGAG 600
 Db 584 GTTACAGCCCGTGTGCTGCTCTGTGCGCTGAGAGGCACTTAATTAATTAACACAGAG 643
 QY 601 ACCCAGTGCATCAGCTGAAGAGACCACTTACTACACAGAGAGTTCACCTGCGCTG 660
 Db 644 ACCCAGTGCATCAGCTGAAGAGACCACTTACTACACAGAGAGTTCACCTGCGCTG 703
 QY 661 AAAGGAATTACGCGCGGAATGACCGAACATGAGAGAGGGGCTGCGGACCCAGTG 720
 Db 704 AAAGGAATTACGCGCGGAATGACCGAACATGAGAGAGGGGCTGCGGACCCAGTG 763
 QY 721 CTCTACCTGCGCGGAGAGTTCACACCGAGTGCCTTGGCGCTGTACACAGTACAC 780
 Db 764 CTCTACCTGCGCGGAGAGTTCACACCGAGTGCCTTGGCGCTGTACACAGTACAC 823
 QY 781 CTGGCGGAGACTACGCTTGCAGCACTGAGTGGTGGCTTGTGCTTGTCTTCTTC 840
 Db 824 CTGGCGGAGACTACGCTTGCAGCACTGAGTGGTGGCTTGTGCTTGTCTTCTTC 883
 QY 841 AACGTGCTGCTTCCAGCCCGCGCCCGCTTACAGGAGGCTTGGAGCTGACACCGGA 900
 Db 884 AACGTGCTGCTTCCAGCCCGCGCCCGCTTACAGGAGGCTTGGAGCTGACACCGGA 943
 QY 901 GCTTCGCGCTTTCGCGGCTTTCGCTTTCGCTTTCATCTTACGCTGCGCTTTCGCG 960
 Db 944 GCTTCGCGCTTTCGCGGCTTTCGCTTTCGCTTTCATCTTACGCTGCGCTTTCGCG 1003
 QY 961 CTTCGCTTACGCTTTCGCGGCTTTCGCTTTCGCTTTCATCTTACGCTGCGCTTTCG 1020
 Db 1004 CTTCGCTTACGCTTTCGCGGCTTTCGCTTTCGCTTTCATCTTACGCTGCGCTTTCG 1063
 QY 1021 GCAACCGGCGCTGCTTTCCTTCTTCCGAGAGGCGCTGTGAGTCTTCCAGTATGTTCCG 1080
 Db 1064 GCAACCGGCGCTGCTTTCCTTCTTCCGAGAGGCGCTGTGAGTCTTCCAGTATGTTCCG 1123
 QY 1081 CCCAGGCTTTCGCACTTTCGCAACCAAGCGCCAGAGCTGCAAGCAGAGAGAGG 1140
 Db 1124 CCCAGGCTTTCGCACTTTCGCAACCAAGCGCCAGAGCTGCAAGCAGAGAGAGG 1183
 QY 1141 GGTCACTCTTATCTTCGCGGACCCACTGCAACAGAGGCGCTTCCAGACTTAAA 1200
 Db 1184 GGTCACTCTTATCTTCGCGGACCCACTGCAACAGAGGCGCTTCCAGACTTAAA 1243
 QY 1201 TGTATACCACTTAAGCTGAGAGGAGACCAATCTGAGCTCTTCCCGCTTGGAGCAT 1260
 Db 1244 TGTATACCACTTAAGCTGAGAGGAGACCAATCTGAGCTCTTCCCGCTTGGAGCAT 1303

QY	126	CGAGGCGCGGAAAGCAAGTGC	CCGCGCAGGCGCTGGGCGACGAGACTC	CAAGAAAGGCACTG	1320
	1304	CGAGGCGCGGAAAGCAAGTGC	CCGCGCAGGCGCTGGGCGACGAGACTC	CAAGAAAGGCACTG	1368
QY	1321	AGCGCTGCTGGCGCGAGGCTT	GGACATCCGCGAGGCAACAGGAAAT	CTCTCTGGGCGGA	1380
Db	1364	AGCGCTGCTGGCGCGAGGCTT	GGACATCCGCGAGGCAACAGGAAAT	CTCTCTGGGCGGA	1428
QY	1381	TCTGTAAATPAAACCTTTT	TTCTTTTGTATTTTAAAAA	AAAAAAAAAAAAAAAAAAAA	1438
1481	1424	TCTGTAAATPAAACCTTTT	TTCTTTTGTATTTTAAAAA	AAAAAAAAAAAAAAAAAAAA	1481

SEQUENCE ID NO 1097; 2081bp + Sequence Listing; English.

The invention relates to novel genes (ABJ89449-ABJ9053) and proteins (ABJ89040-ABJ90444) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant) agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pat_sequences](http://wipo.int/pub/published_pat_sequences)

Sequence 1474 BP; 292 A; 480 C; 404 G; 236 T; 0 U; 2 Other;

Query Match	89.2%	Score 1285;	DB 6;	Length 1474;
Best Local Similarity	99.8%	Pred. No. 0;		
Matches 1435; Conservative	0;	Mismatches	3;	Indels 0; Gaps 0;

QY	1	AAAGTAAAGGGCTCAACGACAGTGAAGAAATATCTTCGTCTGGCCGGAGCTTAAAGAAATCTGTGCG	60
Db	27	AAAGTAAAGGGCTCAACGACAGTGAAGAAATATCTTCGTCTGGCCGGAGCTTAAAGAAATCTGTGCG	86
QY	61	GTACCAACCCCAAGACCGTTGAGAGACGCCACCTCCACGCTTCTTTAAACGAGAGTGTCA	120
Db	87	GTACCAACCCCAAGACCGTTGAGAGACGCCACCTCCACGCTTCTTTAAACGAGAGTGTCA	146
QY	121	GGAACTCAAGCTTCAACGAGCCCACTGGCTCCACGCTTGAAGCAAAAGAGAGCGCCAAAGAC	180
Db	147	GGAACTCAAGCTTCAACGAGCCCACTGGCTCCACGCTTGAAGCAAAAGAGAGCGTCAAGAC	206
QY	181	GGGCTCTCCCGGCTCAAGGACGCCCAAGCTTGTGGCTTGTCTGCCCGCTGCGTGCAGC	240
Db	207	GGGCTCTCCCGGCTCAAGGACGCCCAAGCTTGTGGCTTGTCTGCCCGCTGCGTGCAGC	266
QY	241	ACTCGGCGGGGTGGAGATGATACCTGTGGAAAGGGGTGCTGCTTTTAAACCCCAAGCC	300
Db	267	ACTCGGCGGGGTGGAGATGATACCTGTGGAAAGGGGTGCTGCTTTTAAACCCCAAGCC	326
QY	301	CGGCAATGCGGAGCTTCAAGCTTCACTGCTCATCGTATTTCTAGTGTTTTGTGCTTA	360
Db	327	CGGCAATGCGGAGCTTCAAGCTTCACTGCTCATCGTATTTCTAGTGTTTTGTGCTTA	386
QY	361	GGAGCAAGTTCCTGCTCATCTTTCGCGGGGATCCGTTGGCACTGCGGCGGTGTGTG	420
Db	387	GGAGCAAGTTCCTGCTCATCTTTCGCGGGGATCCGTTGGCACTGCGGCGGTGTGTG	446
QY	421	GTGAGAGTCTTCTCAGTCTGTGTTCAATGAGGCGCAAAATTGTGGCTGTGTGCACTTCA	480
Db	447	GTGAGAGTCTTCTCAGTCTGTGTTCAATGAGGCGCAAAATTGTGGCTGTGTGCACTTCA	506
QY	481	GAATGTTCTGTGGGTGACAGTGAACCAACAACATCTCTAAGAGCCTTCAAGCGGACGCGC	540
Db	507	GAATGTTCTGTGGGTGACAGTGAACCAACAACATCTCTAAGAGCCTTCAAGCGGACGCGC	566
QY	541	GTTACAGCCCGTGTGGTGTGCTGTGGGCTGTGAAGGGCAATTATTAACACTCAACAGG	600
Db	567	GTTACAGCCCGTGTGGTGTGCTGTGGGCTGTGAAGGGCAATTATTAACACTCAACAGG	626
QY	601	ACCCCAAGTGCATCAGCTTAAGAGAACATTTGATCTAACAAGAGATTCACCTGGGGTGTG	660
Db	627	ACCCCAAGTGCATCAGCTTAAGAGAACATTTGATCTAACAAGAGATTCACCTGGGGTGTG	686
QY	661	AAAGAAATTAGCGCGCGAGTACGCGAAGCACTGGAGAAAGGGCTGCGGACCCAGTG	720
Db	687	AAAGAAATTAGCGCGCGAGTACGCGAAGCACTGGAGAAAGGGCTGCGGACCCAGTG	746
QY	721	CTCTACTGTGGGGAATAATTCAACGAGATGAGCCCTTGGGCTGTGACACAAGATACAC	780
Db	747	CTCTACTGTGGGGAATAATTCAACGAGATGAGCCCTTGGGCTGTGACACAAGATACAC	806
QY	781	CTGAGCGAGACATAGCGCTTCGCGCACCTATGAGTGGCTTGTGCTTCTGACTCTCTCC	840
Db	807	CTGAGCGAGACATAGCGCTTCGCGCACCTATGAGTGGCTTGTGCTTCTGACTCTCTCC	866
QY	841	AAAGTGTGTCTTCAAGCGCGGCGCGCTCTAAGAGAGCTGTGCAATGTGTACACACGGA	900
Db	867	AAAGTGTGTCTTCAAGCGCGGCGCGCTCTAAGAGAGCTGTGCAATGTGTACACACGGA	926
QY	901	GACCTTGGCGCTTTCGAGGATCTTTCGCTTGGCTTCATCTTAGCGTGCCTGTGCCG	960
Db	927	GACCTTGGCGCTTTCGAGGATCTTTCGCTTGGCTTCATCTTAGCGTGCCTGTGCCG	986
QY	961	CTTCGGCTTAGGCTCTCCGCGGCTACACAATCAAGTACGAGCGCGCCTTCTGGGCTACGCGT	1020
Db	987	CTTCGGCTTAGGCTCTCCGCGGCTACACAATCAAGTACGAGCGCGCCTTCTGGGCTACGCGT	1046
QY	1021	GCACACGAGGTCTGTGCTCTTCTCGAAGGGCGGTGTGAGTCTCCAGTATGTTCCG	1080

RESULT 5
AAC95564
ID AAC95564 standard; cDNA; 1460 BP.

Human; secreted protein; cytosol; immunosuppressive; nocrotic; neuroprotective; antiviral; anti-allelic; hepatocytic; antidiabetic; anti-inflammatory; antitumor; anticonvulsant; antibacterial; antifungal; antiparasitic; cardiac; cancer; immune disease; allergy; cardiovascular disorder; wound healing; infection; neurological disease.

MO200061596-A1.

06-APR-2000; 2000WO-US008983.

(ROSE/) ROSEN C A.

Ruben SM, Komatsoulis G,

WPI; 2000-611865/58.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

the prevention, treatment and diagnosis of cancer, neurological disorders and neurological diseases

Page 443: English

polymyxin B resistance sequences AAC95521 - AAC95570 represent cDNA encoding 50

AAB52103 represent alternative polypeptides encoded by the genes, and

Query Match	69.1%	Score 996	DB 3	Length 1460
Best Local Similarity	99.9%	Pred. No. 0		
Matches 1046; Conservative	0	Mismatches	1	Indels 0; Gaps 0

QY	AAAGTAAAGCGCTACAGACAGCTGAGAAATATGTTTGGCTGCGCGGCTAGAAAACCTGCTG	60
Db	AAAGTAAAGCGCTACAGACAGCTGAGAAATATGTTTGGCTGCGCGGCTAGAAAACCTGCTG	81
QY	61 GTACCAACCCCAAGAGCGTTGAGAGCAGCCACCTCCACGCTTCCTTAACGGAGAGTGA	120
Db	GTACCAACCCCAAGAGCGTTGAGAGCAGCCACCTCCACGCTTCCTTAACGGAGAGTGA	141
QY	121 GGAAGTCAAGATTGACCAAGCCCACTGGTCCCAAGCCTTGACGAAAGAGACGCGCAAGAC	180
Db	GGAAGTCAAGATTGACCAAGCCCACTGGTCCCAAGCCTTGACGAAAGAGACGCGCAAGAC	201
QY	181 GCGCTCTCCGCGCTCCAGGACGCCCCAGCTTGCTGGCTGGCTGGCCCGCTGGTGGAC	240
Db	GCGCTCTCCGCGCTCCAGGACGCCCCAGCTTGCTGGCTGGCTGGCCCGCTGGTGGAC	261
QY	241 ACTCGCGCGCGCGTGGAGCATGACCGTGGGAAACGGCGATCGCTTTTACCCCAAGCC	300
Db	ACTCGCGCGCGCGTGGAGCATGACCGTGGGAAACGGCGATCGCTTTTACCCCAAGCC	321
QY	301 CGGCAATCGCGAGGCTTACAGCGTTCCACTGCTCATCGTTATCTAGGTTTGGCTCA	360
Db	CGGCAATCGCGAGGCTTACAGCGTTCCACTGCTCATCGTTATCTAGGTTTGGCTCA	381
QY	361 GCAACAGAGTTCTCGTCTCATCTTGCACGGGGAATCCGTGGCCACTGCGCGTGGTGGT	420
Db	GCAACAGAGTTCTCGTCTCATCTTGCACGGGGAATCCGTGGCCACTGCGCGTGGTGGT	441
QY	421 GTGAGAGTTCTTCTCAATCTGTTCATAGCGCGAAGAAATGTGGCTGTGCACTTCAATGA	480
Db	GTGAGAGTTCTTCTCAATCTGTTCATAGCGCGAAGAAATGTGGCTGTGCACTTCAATGA	501
QY	481 GAAATGTTCTGGGTACGTACGTAAACCAACCATCTCAAGAAAGCTTCAGGACGCGCG	540
Db	GAAATGTTCTGGGTACGTACGTAAACCAACCATCTCAAGAAAGCTTCAGGACGCGCG	561
QY	541 GTTACAGACCCGCTGCGGCTGTGCTGTGGGCTCGAGGGGCAATTAATATTAACACTACAGG	600
Db	GTTACAGACCCGCTGCGGCTGTGCTGTGGGCTCGAGGGGCAATTAATATTAACACTACAGG	621
QY	601 ACCCAAGTGCATCAGCTGAAACGACACATTTGACTCAACAGAGCAATTCACTGGCGCTG	660
Db	ACCCCAAGTGCATCAGCTGAAACGACACATTTGACTCAACAGAGCAATTCACTGGCGCTG	681
QY	661 AAAGGAAATTAACGCGCGAGTACGCGAACGACTGAGAAAGGAGGCTCCCGGACCTCAGTG	720

Db	682	AAAGAAATTACCGCGGAGTAAAGCAAGCACTGGAAGAAGGAGCTGCGCGAATCCCATGTG	741
Qy	721	CTCTACCTGGGGGAAABAATTCAACAGATAGACCTTTGGGCGCTTGACACCAAGTACAC	780
Db	742	CTCTACTGGCGGAAGATTACAACCGAGTAGCCCTTGGCGCTTGACACCAAGTACAC	801
Qy	781	CTGGCGGAGACACTACGCGCTCGGACACAGCTATAGGATGGCGCTTGTGCTTGGCTCTCC	840
Db	802	CTGGCGGAGACACTACGCGCTCGGACACAGCTATAGGATGGCGCTTGTGCTTGTCTCTCC	861
Qy	841	AACGTGCTGCTTCCACACCGCGCGCGCTGTAAGGAGCGCTGGACATGTGATCAACACGGA	900
Db	862	AACGTGCTGCTTCCACACCGCGCGCGCTGTAAGGAGCGCTGGACATGTGATCAACACGGA	921
Qy	901	GCGTTGCGGCTCTTTGGGAGTCTTGGCTTGGCTTCATCTTACGCGTGCCTGTGCGCG	960
Db	922	GCGTTGCGGCTCTTTGGGAGTCTTGGCTTGGCTTCATCTTACGCGTGCCTGTGCGCG	981
Qy	961	CTCGGCGTATGAGCGCTCCGCGGCTACCACTAAGTACGAGCGCGCGCTTGGGATCAAGCTG	1020
Db	982	CTTCGCGCTATGAGCGCTCTCCGCGCTCACCACTAAGTACGAGCGCGCGCTTGGGATCAAGCTG	1041
Qy	1021	GCAACCGGCGCTCTGTGCTCTTTCCTC	1047
Db	1042	GCAACCGGCGCTCTGTGCTCTTTCCTC	1068

RESULT 6
AB267235
ID AB267235 standard; cDNA; 1460 BP.
XX
AC AB267235;
XX
DT 26-MAR-2003 (first entry)
XX
DE Human secreted protein encoding cDNA SEQ ID NO 355.
XX
KW Human; secreted protein; neurotropic; neuroprotective; cytosstatic;
KW vitucide; dermatological; immunosuppressive; anti-inflamatory; anti-HIV;
KW vulnerable; antibacterial; antiparkinsonian; antistoking; antianemic;
KW antiatheritic; cancer; antirheumatic; hepatotropic; cerebroprotective;
KW antiinflammatory; antiallergic; antidiabetic; anticancer; anticonvulsant;
KW antifungal; antiparasitic; cardiac; immune disorder; infection; vaccine;
KW cardiovascular disorder; neurological disease; neurotropic;
KW gene therapy; gene; chromosome 9p21; ds.
XX
XX Homo sapiens.
OS
XX
PN W0200277186-A2.
XX
PD 03-OCT-2002.
XX
PF 26-MAR-2002; 2002W0-US009186.
XX
PR 27-MAR-2001; 2001US-0278650P.
XX
PR 12-SEP-2001; 2001US-00950082.
XX
PR 12-SEP-2001; 2001US-00950083.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-040583/03.
XX
DR P-PSDB; ABP99814.
XX
PT New human secreted proteins encoded by genes contained in cDNA clones
PT (e.g. HGCGC19), useful for preventing, treating or diagnosing e.g. AIDS,
PT multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or
PT West Nile fever.
XX
PS Claim 7, Page 1353; 2423pp; English.
XX
CC The invention relates to novel human genes (AB266891-AB268209) and the

CC encoded secreted proteins (ABP99470-0ABP99672) useful for preventing,
CC treating or ameliorating medical conditions e.g. by protein or gene
CC therapy. The genes are isolated from a range of human tissues disclosed
CC in the specification. The nucleic acid, proteins, antibodies and
CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
CC lung or urogenital; (b) immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
CC myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.
CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
CC bacterial, fungal and parasitic infections

XX

SO Sequence 1460 BP; 292 A; 472 C; 398 G; 298 T; 0 U; 0 Other;

QY	1	AAAGTACGGCTACACGACGAGGAAATGTTTCGCTCGCCGGCTGAAAAAATCTGCTCG	50
Db	22	AAAGTACGGCTACACGACGAGGAAATGTTTCGCTCGCCGGCTGAAAAAATCTGCTCG	81
QY	61	GTACCAACCCCGAGACCGTTGAGAGCAGCCCACTCCACGCTTCTTTAACGGAGGTGCA	120
Db	82	GTACCAACCCCGAGACCGTTGAGAGCAGCCCACTCCACGCTTCTTTAACGGAGGTGCA	141
QY	121	GGAGCTAGACTTACACACCCCACTCGCTCCAGCTTGTATGCGAAGAGACGCGCAAGAC	180
Db	142	GGAGCTAGACTTACACACCCCACTCGCTCCAGCTTGTATGCGAAGAGACCTCAAGAC	201
QY	181	GGCGCTTCCCGGCTGCGAGGCAAGCCCACTTGCGCTTGCCTGACCCGCTGCGTACG	240
Db	202	GGCGCTTCCCGGCTGCGAGGCAAGCCCACTTGCGCTTGCCTGACCCGCTGCGTACG	261
QY	241	ACTCGGCGCGGCTGACGATGATCCCTGTGGAAACGGCGCTATGCTTTTAAACCCCAAGCC	300
Db	262	ACTCGGCGCGGCTGACGATGATCCCTGTGGAAACGGCGCTATGCTTTTAAACCCCAAGCC	321
QY	301	CGGCAATGCGCGAGGCTTACAGGTTCCACTGCTCATTCGTTATTCAGTGTGTTTGCTCTA	360
Db	322	CGGCAATGCGCGAGGCTTACAGGTTCCACTGCTCATTCGTTATTCAGTGTGTTTGCTCTA	381
QY	361	GCAGCAAGCTTCTCTCATCTTGCCGGGGAATCCGTGCGCATCGCGGCTGTGTTTGCTG	420
Db	382	GCAGCAAGGTTCTCTCATCTTGCCGGGGAATCCGTGCGCATCGCGGCTGTGTTTGCTG	441
QY	421	GTAGAGAGTCTTCTCAGTCTGTGCTAAGCGCCAAATAATGCGCTGTGCACTTCAGTGA	480
Db	442	GTAGAGAGTCTTCTCAGTCTGTGCTAATGCGCCAAATAATGCGCTGTGCACTTCAGTGA	501
QY	481	GAATGTTGCTGTGGTACAGTGAACCAACCACTCTTCAAAGCTTCAAGCGAGCGCGC	540
Db	502	GAATGTTGCTGTGGTACAGTGAACCAACCACTCTTCAAAGCTTCAAGCGAGCGCGC	561
QY	541	GTTACACGCCGCTGTGCTGCTGCGGCGCTGAGAGGGCAATTAAATTACATCAACAGG	600
Db	562	GTTACACGCCGCTGTGCTGCTGCGGCGCTGAGAGGGCAATTAAATTACATCAACAGG	621
QY	601	ACCCCACTGATCAGCTGAACGAGCACTTGACTTACACGAGCAAGTTCACTTGCGGTCTG	660
Db	622	ACCCCACTGATCAGCTGAACGAGCACTTGACTTACACGAGCAAGTTCACTTGCGGTCTG	681
QY	661	AAAGAGAAATTACGCGCGGAGTACGCGAAACGCACTGGAAGAAAGGGGCTGCGGACCAAGT	720
Db	682	AAAGAGAAATTACGCGCGGAGTACGCGAAACGCACTGGAAGAAAGGGGCTGCGGACCAAGT	741
QY	721	CTTCACTCTGCGAGAGATTCAACACGATGACCTTTCGCGCTGTACCAACAGTACAC	780
Db	742	CTTCACTCTGCGAGAGATTCAACACGATGACCTTTCGCGCTGTACCAACAGTACAC	801

QY 781 CTGGGGGAGACTAGCGCTCGGCGACGCTATGAGGCTTCTGCTTCTGAGCTCTCC 840
 Db 802 CTGGGGGAGACTAGCGCTCGGCGACGCTATGAGGCTTCTGCTTCTGAGCTCTCC 861
 QY 841 AACGTGCTCTCTCCACGCGCGCGCTCTACGAGGCGCTGACGCTGACACCGGA 900
 Db 862 AACGTGCTCTCTCCACGCGCGCGCTCTACGAGGCGCTGACGCTGACACCGGA 921
 QY 901 GCGTTCGCGCTCTTGGGGGTCTTGGCTTGGCTTCGCTTCGCTGAGCGCTGCGCG 960
 Db 922 GCGTTCGCGCTCTTGGGGGTCTTGGCTTGGCTTCGCTTCGCTGAGCGCTGCGCG 981
 QY 961 CTCGCGCTAGGCTCTCTCGGCTCGCTACGACTAGTACGAGCGCGCTTCTGAGTACGCTG 1020
 Db 982 CTCGCGCTAGGCTCTCTCGGCTCGCTACGACTAGTACGAGCGCGCTTCTGAGTACGCTG 1041
 QY 1021 GCACCGGCGCTCTCTGCTCTTCTCTC 1047
 Db 1042 GCACCGGCGCTCTCTGCTCTCTCTC 1068

RESULT 7

AB273640
 ID AB273640 standard; cDNA, 1460 BP.

AC AB273640;

DT 12-MAY-2003 (first entry)

DE Secreted protein-encoding gene 360 cDNA clone HUFCL31. SRQ ID NO:370.

XX Human; secreted protein; cancer; tumour; hyperproliferative disorder;
 XX autoimmune disorder; inflammation; angiogenic diseases; AIDS;
 KM acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
 KM drug screening; chromosome identification; chromosome mapping;
 KM cytoskeletal; gene therapy; antiinflammatory; immunomodulator; anti-HIV;
 KM antihaemic; vulnery; chromosome 9p21; gene; ss.

XX Homo sapiens.

XX WO200277013-A2.

XX 03-OCT-2002.

XX 26-MAR-2002; 2002MO-US009370.

XX 27-MAR-2001; 2001US-0278650P.

XX 12-SEP-2001; 2001US-00950082.

XX 12-SEP-2001; 2001US-00950083.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2003-040578/03.

XX P-PSDB; ABR01306.

XX New human secreted proteins and nucleic acids, useful for detecting or

XX treating cancer or other hyperproliferative disorders, autoimmune

XX disorders, inflammatory disorders, HIV disease, hepatitis or anemia.

XX Claim 21; Page 1345; 24749P; English.

XX AB273281-AB273697 represent cDNAs corresponding to 391 human secreted
 CC protein genes, and ABP00947-ABP01363 represent the proteins they encode.
 CC AB273698-AB274687 represent human secreted protein genomic fragments. The
 CC invention also encompasses antibodies specific for the secreted proteins,
 CC the use of the secreted proteins in drug screening and recombinant
 CC vectors and host cells comprising a nucleic acid of the invention. The
 CC secreted proteins are thought to be involved in biological activities
 CC associated with cellular signalling, cellular differentiation, cell
 CC migration, prohormone activation and neurotransmitter activity. The
 CC secreted proteins, nucleic acids encoding them, antibodies or antibody

CC fragments specific for the secreted proteins, and modulators of protein
 CC activity are useful for diagnosing or treating cancers or other
 CC hyperproliferative disorders. Additionally, the secreted proteins and
 CC their nucleic acids may also be used in the treatment of autoimmune
 CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
 CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
 CC wound healing. Nucleic acids of the invention may be used for chromosome
 CC identification, chromosome mapping, in gene therapy, for identifying
 CC individuals from minute biological samples, as hybridisation probes, and
 CC as molecular weight markers. The present sequence represents a human
 CC secreted protein-encoding cDNA clone of the invention

XX SQ Sequence 1460 BP; 292 A; 472 C; 398 G; 298 T; 0 U; 0 Other;

XX Query Match 69.1%; Score 996; DB 7; Length 1460;

XX Best Local Similarity 99.9%; Pred. No. 0;

XX Matches 1046; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAGTAACGCTACACAGACGAGTAATAGTTTGGCTCGCGGCTGAAAAAATCTGTG 60

Db 22 AAAGTAACGCTACACAGACGAGTAATAGTTTGGCTCGCGGCTGAAAAAATCTGTG 81

QY 61 GTACCAACCCGAGAGCTTGAAGACAGCCACCTCCAGCTTCTTAACGAGAGTGA 120

Db 82 GTACCAACCCGAGAGCTTGAAGACAGCCACCTCCAGCTTCTTAACGAGAGTGA 141

QY 121 GGAATCAGACTTTCACAGAGCCACTGCGTCCGACCTTGTACGAAAGAGAGGAC 180

Db 142 GGAATCAGACTTTCACAGAGCCACTGCGTCCGACCTTGTACGAAAGAGAGGAC 201

QY 181 GGGCTCTCCGCGGTCCAGAGGAGCCAGCCAGCTTGTGAGCTTGTGCGCGCTGAGT 240

Db 202 GGGCTCTCCGCGGTCCAGAGGAGCCAGCCAGCTTGTGAGCTTGTGCGCGCTGAGT 261

QY 241 ACTCGCGCGCGGTGACAGATGACCTGTGAGACGCGGTACTGCTTTTACCCCGAG 300

Db 262 ACTCGCGCGCGGTGACAGATGACCTGTGAGACGCGGTACTGCTTTTACCCCGAG 321

QY 301 CGGCATGCGCGAGAGCTTGAAGCTGCACTGCTATGCTTATGAGTGTGTTGGCTCT 360

Db 322 CGGCATGCGCGAGAGCTTGAAGCTGCACTGCTATGCTTATGAGTGTGTTGGCTCT 381

QY 361 GCAGCAAGCTTCTGCTCATCTTTCGCGGAGATCCGTGAGCACTGCGCTGTTTGTG 420

Db 382 GCAGCAAGCTTCTGCTCATCTTTCGCGGAGATCCGTGAGCACTGCGCTGTTTGTG 441

QY 421 GTGAGATTTCTTCTAGTCTGTTCATAGAGCGCAAAATGTGCTGTGCACTTCA 480

Db 442 GTGAGATTTCTTCTAGTCTGTTCATAGAGCGCAAAATGTGCTGTGCACTTCA 501

QY 481 GAATGATGTTGGGTATAGTGAACCAACACATCTTCAAAAGCTTCAGCGACGCGC 540

Db 502 GAATGATGTTGGGTATAGTGAACCAACACATCTTCAAAAGCTTCAGCGACGCGC 561

QY 541 GTTACAGCCCGTGTGCTGTGCTGTGAGGCTTGAAGGCAATTAATTTTACTACAG 600

Db 562 GTTACAGCCCGTGTGCTGTGCTGTGAGGCTTGAAGGCAATTAATTTTACTACAG 621

QY 601 ACCCGAGTGCATCAGTGAACGAGACATTTGATTAACAGAGCTTCACTGCGCTG 660

Db 622 ACCCGAGTGCATCAGTGAACGAGACATTTGATTAACAGAGCTTCACTGCGCTG 681

QY 661 AAAGAGAAATTAAGCGCGAGAGTACGCGAAGCGACTGGAAGGGGCTGCGGACCG 720

Db 682 AAAGAGAAATTAAGCGCGAGAGTACGCGAAGCGACTGGAAGGGGCTGCGGACCG 741

QY 721 CTTTACTGCGGAGAGTTCACACGAGTACCTTTCGCGCTGTACCAACGATACAC 780

Db 742 CTTTACTGCGGAGAGTTCACACGAGTACCTTTCGCGCTGTACCAACGATACAC 801

QY 781 CTGGCGGAGACATTAAGCGCTGCGACAGCTATGAGTGTGCTTCTGCTCTCTCC 840

Db 802 CTGGCGGAGACATTAAGCGCTGCGACAGCTATGAGTGTGCTTCTGCTCTCTCTCC 861

QY	841	AAAGTGTCTCTCTCCACGCGGAGCCCGGCTCTACGGAGAGCCTTGGCACTTGTATACACCGGA	900
Db	862	AAAGTGTCTCTCTCCACGCGGAGCCCGGCTCTACGGAGAGCCTTGGCACTGTATACACCGGA	921
QY	901	GCCTTGAGCCTCTTTCGGGGTCTTGACCTTGGCTCCATCTCTAGACGTGCCTGTGCCG	960
Db	922	GCCTTGAGCCTCTTTCGGGGTCTTGACCTTGGCTCCATCTCTAGACGTGCCTGTGCCG	981
QY	961	CTTCGGCTTAGGCTCTTCGCGGCTCAACAATTAGTACGGGCGCGGCTTCTGGGTACAGCTG	1020
Db	982	CTTCGGCTTAGGCTCTTCGCGGCTCAACAATTAGTACGGGCGCGGCTTCTGGGTACAGCTG	1041
QY	1021	GCACCGGCGCTCTGTGGCTCTTCTCTC	1047
Db	1042	GCACCGGCGCTCTGTGGCTCTTCTCTC	1068

RESULT 8	
ADCC20289	
ID	ADCC20289 standard; DNA; 1460 BP.
AC	
XX	ADCC20289;
XX	
DT	18-DEC-2003 (first entry)
XX	
DE	Human secreted protein coding sequence #228.
XX	
KW	gene therapy; human; secreted protein; haemopoietic disorder;
KW	haematological disorder; anaemia; haemophilia; inflammatory disorder;
KW	inflammatory bowel disease; Crohn's disease; neoplastic disease; cancer;
KW	leukaemia; wound healing; epithelial cell proliferation disorder;
KW	immune disorder; autoimmune disorder; asthmatic disorder;
KW	cardiovascular disorder; atherosclerosis; myocarditis;
KW	infectious disease; HIV; AIDS; endocrine disorder; diabetes;
KW	gastrointestinal disorder; duodenal ulcer; gastroenteritis; gene; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200292787-A2.
XX	
PD	21-NOV-2002.
XX	
PF	26-MAR-2002; 2002MO-US009257.
XX	
PR	27-MAR-2001; 2001US-0278650P.
PR	12-SEP-2001; 2001US-00950082.
PR	12-SEP-2001; 2001US-00950083.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Rosen CA, Ruben SM;
XX	
DR	WPI; 2003-129287/12.
XX	
PT	New human secreted proteins and nucleic acid molecules, useful for
PT	preparing a diagnostic or pharmaceutical composition for diagnosing,
PT	preventing or treating hematopoietic or hematologic disorders, e.g.
PT	anemia or hemophilia.
XX	
PS	Claim 1; SEQ ID NO 238; 1512bp; English.
XX	
CC	The invention comprises the amino acid and coding sequences of human
CC	secreted proteins. The DNA and protein sequences of the invention are
CC	useful for detecting, preventing, diagnosing, prognosticating, treating
CC	or ameliorating: haematopoietic or haematological disorders (e.g. anaemia
CC	and haemophilia); inflammatory disorders (e.g. inflammatory bowel disease
CC	and Crohn's disease); neoplastic disease (e.g. cancer and leukaemia);
CC	wound healing and disorders of epithelial cell proliferation; immune
CC	disorders (e.g. autoimmune disorders and asthmatic disorders);
CC	cardiovascular disorders (e.g. atherosclerosis and myocarditis);
CC	infectious disease (e.g. HIV/AIDS); endocrine disorders (e.g. diabetes);
CC	and gastrointestinal disorders (e.g. duodenal ulcers and

CC gastroenteritis). The present DNA sequence encodes a human secreted
 CC protein of the invention.
 CC
 XX
 SQ Sequence 1460 BP; 292 A; 472 C; 398 G; 298 T; 0 U; 0 Other;

 Query Match 69.1%; Score 996; DB 9; Length 1460;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1046; Conservative 0; Mismatches 1; Indels 0; Gaps 0

QY	1	AAAGTAA	CGGCTAA	CAGACAG	AGAGAA	TAAGTTTG	CGCTCG	CGCGCTAA	AAAAAC	CTGTCG	60
Db	22	AAAGTAA	CGGCTAA	CAGACAG	TAAGTTTA	AGTTTG	CGCTCG	CGCGCTAA	AAAAAC	CTGTCG	81
QY	61	GTACCA	CCCCCAG	AGCGTTG	AAGACG	AGCCCA	CTTCA	CGCCTT	CCTTAA	CGAAGGTGCA	120
Db	82	GTACCA	AAACCCCAG	AGCGTTG	AAGACG	AGCCCA	CTTCA	CGCCTT	CCTTAA	CGAAGGTGCA	141
QY	121	GGACTCA	GACTTAC	CCAGCCCA	CTGTGGTCC	CAAGCCTT	GTAGCG	AAAGAGCG	CAAGGAC	180	
Db	142	GGACTCA	GACTTAC	CCAGCCCA	CTGTGGTCC	CAAGCCTT	GTAGCG	AAAGAGCG	CAAGGAC	201	
QY	181	GGCCTCT	CCCGCG	CTCCAGG	CAGCCCA	CTTGCTGTG	CGTCCG	CGCTGCGT	CGACG	240	
Db	202	GGCCTCT	CCCGCGT	CCAGGCA	CCCACTTG	CGTGGCTG	CGTCCG	CGCTGCGT	CGACG	261	
QY	241	ACTCGCG	CGCGCTG	CAGCATG	ACCTGTG	GAACGCGT	ACTGCTTTT	TAACCCCA	CGCC	300	
Db	262	ACTCGCG	CGCGCTG	CAGCATG	ACCTGTG	GAACGCGT	ACTGCTTTT	TAACCCCA	CGCC	321	
QY	301	CGGATAC	CGCAGCGTT	CAGCGTT	CACTGCAT	CGTATTT	CAGTGT	TTTGGCTTA	360		
Db	322	CGGATAC	CGCAGCGTT	CAGCGTT	CACTGCAT	CGTATTT	CAGTGT	TTTGGCTTA	381		
QY	361	GCAGCA	AGCTTCT	CTCATCTT	CGCGG	GAATCCG	TGSCA	CTCGCGCT	TGGTTG	420	
Db	382	GCAGCA	AGCTTCT	CTCATCTT	CGCGG	GAATCCG	TGSCA	CTCGCGCT	TGGTTG	441	
QY	421	GTAGAG	AGTTCTT	CTCAGTCT	GTTCAT	PAGCGAG	AAATGTTG	CGCTG	GCAC	480	
Db	442	GTAGAG	AGTTCTT	CTCAGTCT	GTTCAT	PAGCGAG	AAATGTTG	CGCTG	GCAC	501	
QY	481	GAATG	GTTCGTG	GTACAG	TGAACCA	CAACAT	CTTCA	CAAA	AGCTTT	CAAGCG	540
Db	502	GAATG	GTTCGTG	GTACAG	TGAACCA	CAACAT	CTTCA	CAAA	AGCTTT	CAAGCG	561
QY	541	GTTHA	CGCCGTG	CTCGTCTG	CTCGTGG	CGCTG	GAAGG	GCATTA	TATTA	CACTCA	600
Db	562	GTTHA	CGCCGTG	CTCGTCTG	CTCGTGG	CGCTG	GAAGG	GCATTA	TATTA	CACTCA	621
QY	601	ACCCAG	TGATCAG	CTGAA	ACGACG	ACATTTG	CTA	CAACG	AGTTCA	CCTGG	660
Db	622	ACCCAG	TGATCAG	CTGAA	ACGACG	ACATTTG	CTA	CAACG	AGTTCA	CCTGG	681
QY	661	AAAGA	GAATTA	CCGCGCG	GAATTA	CCGCG	GAATTA	CCGCG	GAATTA	CCGCG	720
Db	682	AAAGA	GAATTA	CCGCGCG	GAATTA	CCGCG	GAATTA	CCGCG	GAATTA	CCGCG	741
QY	721	CTTCA	CTGCG	GAAGATT	CA	CCGAGT	AGCCTT	CGGCTGT	CA	CAAGTAC	780
Db	742	CTTCA	CTGCG	GAAGATT	CA	CCGAGT	AGCCTT	CGGCTGT	CA	CAAGTAC	801
QY	781	CTGG	CGGGA	CACTA	CGCTTC	CGGCA	CGTAT	TGGTGG	CTTCTG	CTCTCC	840
Db	802	CTGG	CGGGA	CACTA	CGCTTC	CGGCA	CGTAT	TGGTGG	CTTCTG	CTCTCC	861
QY	841	AAAGT	GTGCT	CTTCA	CGCGCG	CGCGCT	CTTCA	CGGAG	CGCTT	GGACAC	900
Db	862	AAAGT	GTGCT	CTTCA	CGCGCG	CGCGCT	CTTCA	CGGAG	CGCTT	GGACAC	921
QY	901	GCGTT	CGCGCT	CTTCTG	CGGCTT	TGCGCTT	GCATCT	TAAG	CGTTC	CGCTT	960
Db	922	GCGTT	CGCGCT	CTTCTG	CGGCTT	TGCGCTT	GCATCT	TAAG	CGTTC	CGCTT	981

QY 961 CTCGCCCTAGGCTCTCCGCGCTACCACTAGTACGGCGCGCTTCTGAGTACGCTG 1020
 DB 982 CTCGCCCTAGGCTCTCCGCGCTACCACTAGTACGGCGCGCTTCTGAGTACGCTG 1041
 QY 1021 GCAACCGAGGCTCTGCTCTCTCTC 1047
 DB 1042 GCACCGCGCTCTGCTCTCTCTC 1068

RESULT 9

AAA94624
 ID AAA94624 standard; DNA; 498 BP.

AC AAA94624;

DT 11-JAN-2001 (first entry)

DE Human CASB618 EST.

KM Epitope; human; CASB618; chromosome 15; tumour; vaccine; cancer; ovarian;
 KM color; autoimmune disease; HLA_A0201; expressed sequence tag; EST; ss.

OS Homo sapiens.

PN WO200053748-A2.

PD 14-SEP-2000.

PF 09-MAR-2000; 2000WO-EP002048.

PR 11-MAR-1999; 99GB-000050607.

PR 01-SEP-1999; 99GB-00020590.

PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

PI Bruck CEM, Cassart J, Coche T, Vinals Y De Basolsc;

DR WPI; 2000-572268/53.

PT New human CASB618 polypeptide, useful as a vaccine for prophylactic and
 PT therapeutic treatment of cancers, particularly ovarian or colon cancer,
 PT autoimmune diseases and related conditions.

PS Claim 32; Page 62; 76pp; English.

XX The present sequence is an expressed sequence tag (EST) for human CASB618

CC protein. The gene for human CASB618 is thought to be located on

CC chromosome 15. CASB618 protein and epitopes of CASB618 protein (see

CC AAB26327 to AAB26399) are useful in diagnosing the occurrence of tumour

CC cells and in vaccines for prophylactic and therapeutic treatment of

CC cancers, particularly ovarian or colon cancer, autoimmune diseases and

CC related conditions

XX Sequence 498 BP; 97 A; 166 C; 137 G; 98 T; 0 U; 0 Other;

Query Match 29.5%; Score 425; DB 3; Length 498;

Best Local Similarity 100.0%; Pred No. 1.9e-153;

Matches 425; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 998 GCGCGCGCTTCTGAGTACGCTTGCAACCGCGCTGCTCTTCTCTCGAGGCGCG 1057

DB 62 GCGCGCGCTTCTGAGTACGCTTGCAACCGCGCTGCTCTTCTCTCGAGGCGCG 121

QY 1058 TGGTGAAGTCCAGTATGTTTGGGCGGCTCTTGAGCAACCTTGAGCAACAGCGCA 1117

DB 122 TGGTGAAGTCCAGTATGTTTGGGCGGCTCTTGAGCAACCTTGAGCAACAGCGCA 181

QY 1118 AGGACTGACAGCAGAGAGAGAGGAGGCTCACTCTTATCTCGGCAACCACTGCAAGC 1177

DB 182 AGGACTGACAGCAGAGAGAGAGGAGGCTCACTCTTATCTCGGCAACCACTGCAAGC 241

QY 1178 AGGCGGCTTCCAGACTTAATGATACCACTTAAGTGAAGGAGGAGCCCAATCTGG 1237

DB 242 AGCCCGCTCTCCAGACTTAATGATACCACTTAAGTGAAGGAGGAGCCCAATCTGG 301
 QY 1238 ACTCTTCCCGGCTTGGGCAATCGAGAGCGGGGAAGCATGTGCGCCGAGCTGGGCA 1297
 DB 302 ACTCTTCCCGGCTTGGGCAATCGAGAGCGGGGAAGCATGTGCGCCGAGCTGGGCA 361
 QY 1298 GGAAGCTCCAGAGAGGCACTAGAGCGCTGTCGCGCGAGGCTGGAGCATCGGAGGCA 1357
 DB 362 GGAAGCTCCAGAGAGGCACTAGAGCGCTGTCGCGCGAGGCTGGAGCATCGGAGGCA 421
 QY 1358 CCAGGAAAGTCTCTCTGCGGCGATCTGTAATTAACCTTTTCTTTTGTAAAA 1417
 DB 422 CCAGGAAAGTCTCTCTGCGGCGATCTGTAATTAACCTTTTCTTTTGTAAAA 481
 QY 1418 AAAA 1422
 DB 482 AAAA 486

RESULT 10

ADD19187
 ID ADD19187 standard; CDNA; 580 BP.

AC ADD19187;

DT 15-JAN-2004 (first entry)

DE Human CDNA from secreted protein gene 4.

KM human secreted protein; cytosolic; antibacterial; virulence;

KM neuroprotective; gynaecological; gastrointestinal; gen; cardiac;

KM cardiovascular; gen; nephrotropic; anti-inflammatory; muscular; gen;

KM respiratory; gen; immunosuppressive; cerebroprotective; vasotropic;

KM neotropic; anti-allergic; cancer; bacterial infection; viral infection;

KM neural disorder; immune system disorder; blood disorder;

KM muscular disorder; reproductive disorder; gastrointestinal disorder;

KM pulmonary disorder; cardiovascular disorder; renal disorder;

KM inflammatory disorder; proliferative disorder; human; ss; gene.

XX Homo sapiens.

OS WO2003052377-A2.

PN 26-JUN-2003.

PF 06-NOV-2002; 2002WO-US035606.

PR 07-NOV-2001; 2001US-0331046P.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM,

DR WPI; 2003-533050/50.

DR P-PSDB; ADD19262.

PT New isolated nucleic acids encoding signal transduction pathway component

PT polypeptides, useful for diagnosing, treating, and/or preventing

PT diseases, such as cancer, infections, cardiovascular and inflammatory

XX Claim 1; SEQ ID NO 14; 554pp; English.

XX The invention relates to an isolated nucleic acid molecule (cdna)

CC encoding a human secreted protein, representing one of 85 novel genes.

CC Also included are recombinant vectors, host cells (expressing the

CC protein), the secreted proteins (including their fragments, epitopes and

CC homologues), an isolated antibody that binds specifically to the protein,

CC diagnosing a pathological condition or susceptibility to a pathological

CC condition (comprising determining the presence or absence of a mutation

CC in the nucleic acid and diagnosing a condition based on the presence or

CC absence of the mutation), diagnosing a pathological condition or

CC susceptibility to a pathological condition (comprising determining the

CC presence or amount of expression of the protein in a biological sample
 CC and diagnosing a condition based on the presence or amount of expression
 CC of the protein), preventing, treating or ameliorating a medical condition
 CC by administering the nucleic acid or protein to a mammalian subject,
 CC identifying a binding partner to the protein, the gene corresponding to
 CC the cDNA sequence, and identifying an activity in a biological assay
 CC (comprising expressing the nucleic acid in a cell, isolating the
 CC supernatant, detecting an activity in a biological assay and identifying
 CC the protein in the supernatant having the activity). The nucleic acids
 CC and proteins display the following activities: cytostatic, antibacterial,
 CC virucide, Neuroprotective, Gynaecological, Gastrointestinal-Gen,
 CC Cardiac, Cardiovascular-Gen, Nephrotoxic, Antiinflammatory-Gen,
 CC Respiratory-Gen, Immunosuppressive, Cerebroprotective, Vascular,
 CC Neurotoxic, Antiallergic. The methods and compositions of the present
 CC invention are useful for diagnosing, treating, preventing and/or
 CC prognosticating disorders related to the novel polypeptides, such as
 CC cancer, bacterial or viral infections, and neural, immune system, blood,
 CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
 CC renal, inflammatory or proliferative disorders (many examples of these
 CC diseases and disorders are given in the specification). The present
 CC sequence encodes a novel secreted protein of the invention.

XX Sequence 580 BP; 115 A; 198 C; 156 G; 111 T; 0 U; 0 Other;

Query Match 28.1%; Score 405; DB 9; Length 580;
 Best Local Similarity 100.0%; Pred. No. 8.6e-146;
 Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAGTAAAGGCTACAGACAGTGAAGAAATAGTTTGGCTCCGCGCTGAGAAAACCTGTGCG 60
 DB 13 AAAGTAAAGGCTACAGACAGTGAAGAAATAGTTTGGCTCCGCGCTGAGAAAACCTGTGCG 72
 QY 61 GTACCAACCCAGAGCGTTGAGACAGACCCACCTCCACGCTTCCTTAACGAGAGTGCA 120
 DB 73 GTACCAACCCAGAGCGTTGAGACAGACCCACCTCCACGCTTCCTTAACGAGAGTGCA 132
 QY 121 GGAATCAGACTTTCACAGCCCACTCGGTCCCAAGCTTTGACGAAAGAACCCCAAGAC 180
 DB 133 GGAATCAGACTTTCACAGCCCACTCGGTCCCAAGCTTTGACGAAAGAACCCCAAGAC 192
 QY 181 GCGCTCCCGCGTCCAG 240
 DB 193 GCGCTCCCGCGTCCAG 252
 QY 241 ACTGAGCGCGGTGACAGATGACCTGTGAAAGGCGTACTGCTTTTAAACCCAGAGCC 300
 DB 253 ACTGAGCGCGGTGACAGATGACCTGTGAAAGGCGTACTGCTTTTAAACCCAGAGCC 312
 QY 301 CGGATGCGCGAGGCTTCAGAGGTTCCACTGCTCATCGTATTCAGAGTTTGGCTCTTA 360
 DB 313 CGGATGCGCGAGGCTTCAGAGGTTCCACTGCTCATCGTATTCAGAGTTTGGCTCTTA 372
 QY 361 GCAGCAGCTTCTGCTCATCTTCCGCGGAGATCCGTGGCACTCG 405
 DB 373 GCAGCAGCTTCTGCTCATCTTCCGCGGAGATCCGTGGCACTCG 417

RESULT 11

ABZ68115
 ID ABZ68115 standard; DNA; 5033 BP.

XX ABZ68115;

DT 26-MAR-2003 (first entry)

DE Human secreted protein encoding genomic DNA SEQ ID NO 1638.

XX Human; secreted protein; neurotropic; neuroprotective; cytostatic;
 KW virucide; dermatological; immunosuppressive; antiinflammatory; anti-HIV;
 KW vulnereary; antibacterial; antiparkinsonian; antidiabetic; antianemic;
 KW antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective;
 KW antiinflammatory; antiallergic; antidiabetic; antitumor; anticonvulsant;
 KW antifungal; antiparasitic; cardiant; immune disorder; infection; vaccine;

KW cardiovascular disorder; neurological disease; nephrotoxic;
 KW gene therapy; gene; ds.

XX Homo sapiens.

XX MO20027186-A2.

XX 03-OCT-2002.

XX 26-MAR-2002; 2002MO-US009188.

XX 27-MAR-2001; 2001US-0278650P.

XX 12-SEP-2001; 2001US-00950082.

XX 12-SEP-2001; 2001US-00950083.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2003-040583/03.

XX New human secreted proteins encoded by genes contained in cDNA clones

XX (e.g. HGCAC19), useful for preventing, treating or diagnosed e.g. AIDS,

XX multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or

XX West Nile fever.

XX Disclosure; Page 2263-2264; 2423pp; English.

CC The invention relates to novel human genes (ABZ68891-ABZ68209) and the
 CC encoded secreted proteins (ABP9470-ABP9872) useful for preventing,
 CC treating or ameliorating medical conditions e.g. by protein or gene
 CC therapy. The genes are isolated from a range of human tissues disclosed
 CC in the specification. The nucleic acids, proteins, antibodies and
 CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
 CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the
 CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
 CC lung or urogenital; (b) immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
 CC myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.
 CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
 CC bacterial, fungal and parasitic infections

SO Sequence 5033 BP; 1061 A; 1409 C; 1465 G; 1098 T; 0 U; 0 Other;

Query Match 28.1%; Score 405; DB 7; Length 5033;
 Best Local Similarity 100.0%; Pred. No. 6e-146;
 Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAGTAAAGGCTACAGACAGTGAAGAAATAGTTTGGCTCCGCGCTGAGAAAACCTGTGCG 60
 DB 13 AAAGTAAAGGCTACAGACAGTGAAGAAATAGTTTGGCTCCGCGCTGAGAAAACCTGTGCG 72
 QY 61 GTACCAACCCAGAGCGTTGAGACAGACCCACCTCCACGCTTCCTTAACGAGAGTGCA 120
 DB 73 GTACCAACCCAGAGCGTTGAGACAGACCCACCTCCACGCTTCCTTAACGAGAGTGCA 132
 QY 121 GGAATCAGACTTTCACAGCCCACTCGGTCCCAAGCTTTGACGAAAGAACCCCAAGAC 180
 DB 133 GGAATCAGACTTTCACAGCCCACTCGGTCCCAAGCTTTGACGAAAGAACCCCAAGAC 192
 QY 181 GCGCTCCCGCGTCCAG 240
 DB 193 GCGCTCCCGCGTCCAG 252
 QY 241 ACTGAGCGCGGTGACAGATGACCTGTGAAAGGCGTACTGCTTTTAAACCCAGAGCC 300
 DB 253 ACTGAGCGCGGTGACAGATGACCTGTGAAAGGCGTACTGCTTTTAAACCCAGAGCC 312
 QY 301 CGGATGCGCGAGGCTTCAGAGGTTCCACTGCTCATCGTATTCAGAGTTTGGCTCTTA 360
 DB 313 CGGATGCGCGAGGCTTCAGAGGTTCCACTGCTCATCGTATTCAGAGTTTGGCTCTTA 372

QY 361 GCAGCAAGCTTCTGCTCATCTTGGCCGGGATCCGTGCACTCG 405
 DB 373 GCAGCAAGCTTCTGCTCATCTTGGCCGGGATCCGTGCACTCG 417

RESULT 12
 ID AB274587 standard; DNA; 5033 BP.
 AC AB274587;
 XX
 DT 12-MAY-2003 (first entry)
 XX

DE Secreted protein gene 360 genomic fragment HUPCL31, SEQ ID NO:1734.
 KW Human; secreted protein; cancer; tumour; hyperproliferative disorder;
 KW autoimmune disorder; inflammation; angiogenic diseases; AIDS;
 KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
 KW drug screening; chromosome identification; chromosome mapping;
 KW cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;
 KW antianaemic; vulnery; chromosome 9p21; gene; ds.
 OS Homo sapiens.
 XX
 PN WO200277013-A2.
 PD 03-OCT-2002.
 PF 26-MAR-2002; 2002WO-US009370.
 XX
 PR 27-MAR-2001; 2001US-0278650P.
 PR 12-SEP-2001; 2001US-00950082.
 PR 12-SEP-2001; 2001US-00950083.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2003-040578/03.
 XX
 PT New human secreted proteins and nucleic acids, useful for detecting or
 PT treating cancer or other hyperproliferative disorders, autoimmune
 PT disorders, inflammatory disorders, HIV disease, hepatitis or anaemia.
 XX
 PA Disclosure, Page 2315-2316; 2474pp; English.
 XX
 CC AB273281-AB273697 represent cDNAs corresponding to 391 human secreted
 CC protein genes, and ABP00947-ABP01363 represent the proteins they encode.
 CC AB273698-AB274687 represent human secreted protein genomic fragments. The
 CC invention also encompasses antibodies specific for the secreted proteins,
 CC the use of the secreted proteins in drug screening and recombinant
 CC vectors and host cells comprising a nucleic acid of the invention. The
 CC secreted proteins are thought to be involved in biological activities
 CC associated with cellular signalling, cellular differentiation, cell
 CC migration, pro-hormone activation and neurotransmitter activity. The
 CC secreted proteins, nucleic acids encoding them, antibodies or antibody
 CC fragments specific for the secreted proteins, and modulators of protein
 CC activity are useful for diagnosing or treating cancers or other
 CC hyperproliferative disorders. Additionally, the secreted proteins and
 CC their nucleic acids may also be used in the treatment of autoimmune
 CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS-
 CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
 CC wound healing. Nucleic acids of the invention may be used for chromosome
 CC identification, chromosome mapping, in gene therapy, for identifying
 CC individuals from minute biological samples, as hybridisation probes, and
 CC as molecular weight markers. The present sequence represents a human
 CC secreted protein genomic fragment referred to in the disclosure of the
 CC invention.
 CC
 CC Sequence 5033 BP; 1061 A; 1409 C; 1465 G; 1098 T; 0 U; 0 Other;
 CC
 CC Query Match 28.1%; Score 405; DB 7; Length 5033;

Best Local Similarity 100.0%; Pred. No. 6e-146;
 Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAGTACGGGCTACAGACAGTGAATGTTTGGCTGGCGGGCTAGAAAACCTGTCG 60
 DB 13 AAAGTACGGGCTACAGACAGTGAATGTTTGGCTGGCGGGCTAGAAAACCTGTCG 72
 QY 61 GTACCAACCCCAAGAGGTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
 DB 73 GTACCAACCCCAAGAGGTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 132
 QY 121 GGAATCAAGCTTACAG 180
 DB 133 GGAATCAAGCTTACAG 192
 QY 181 GGCTCTCCGCGCTCAG 240
 DB 193 GGCTCTCCGCGCTCAG 252
 QY 241 ACTGCGCGCGGTGACAGATGACCTGTGTGAACGGGCTACTGCTTTTACCCCGAGCC 300
 DB 253 ACTGCGCGCGGTGACAGATGACCTGTGTGAACGGGCTACTGCTTTTACCCCGAGCC 312
 QY 301 CGGATGCGCGAGGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
 DB 313 CGGATGCGCGAGGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 372
 QY 361 GCAGCAAGCTTCTGCTCATCTTGGCCGGGATCCGTGCACTCG 405
 DB 373 GCAGCAAGCTTCTGCTCATCTTGGCCGGGATCCGTGCACTCG 417

RESULT 13
 ID ADC21005 standard; DNA; 5033 BP.
 AC ADC21005;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE Human secreted protein-related DNA sequence #423.
 XX
 DE gene therapy; human; secreted protein; haemopoietic disorder;
 KW haematological disorder; anaemia; haemophilia; inflammatory disorder;
 KW inflammatory bowel disease; Crohn's disease; neoplastic disease; cancer;
 KW leukaemia; wound healing; epithelial cell proliferation disorder;
 KW immune disorder; autoimmune disorder; asthmatic disorder;
 KW cardiovascular disorder; atherosclerosis; myocarditis;
 KW infectious disease; HIV; AIDS; endocrine disorder; diabetes;
 KW gastrointestinal disorder; duodenal ulcer; gastroenteritis; gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200292787-A2.
 PD 21-NOV-2002.
 PF 26-MAR-2002; 2002WO-US009257.
 XX
 PR 27-MAR-2001; 2001US-0278650P.
 PR 12-SEP-2001; 2001US-00950082.
 PR 12-SEP-2001; 2001US-00950083.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2003-129287/12.
 XX
 PT New human secreted proteins and nucleic acid molecules, useful for
 PT preparing a diagnostic or pharmaceutical composition for diagnosing,
 PT preventing or treating hematopoietic or hematologic disorders, e.g.
 PT anemia or hemophilia.

XX Disclosure; SEQ ID NO 959; 1512pp; English.
PS
XX The invention comprises the amino acid and coding sequences of human
CC secreted proteins. The DNA and protein sequences of the invention are
CC useful for detecting, preventing, diagnosing, prognosticating, treating
CC or ameliorating: haematopoietic or haematological disorders (e.g. anaemia
CC and haemophilia); inflammatory disorders (e.g. inflammatory bowel disease
CC and Crohn's disease); neoplastic disease (e.g. cancer and leukaemia);
CC wound healing and disorders of epithelial cell proliferation; immune
CC disorders (e.g. autoimmune disorders and asthmatic disorders);
CC cardiovascular disorders (e.g. atherosclerosis and myocarditis);
CC infectious disease (e.g. HIV/AIDS); endocrine disorders (e.g. diabetes);
CC and gastrointestinal disorders (e.g. duodenal ulcers and
CC gastroenteritis). The present DNA sequence was used in the
CC exemplification of the invention.
XX
SQ Sequence 5033 BP; 1061 A; 1409 C; 1465 G; 1098 T; 0 U; 0 Other;
Query Match 28.1%; Score 405; DB 9; Length 5033;
Best Local Similarity 100.0%; Pred. No. 66-146;
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAAGTAAAGGCTACAGACAGTGAAGAAATGTTGCTGCGCGGCTAGAAAACTCTGTG 60
Db 13 AAAGTAAAGGCTACAGACAGTGAAGAAATGTTGCTGCGCGGCTAGAAAACTCTGTG 72
QY 61 GTACCAACCCAGAGCGCTTGAAGAGAGCCACCTCCAGCGCTTCTTAAGAGAGGTGA 120
Db 73 GTACCAACCCAGAGCGCTTGAAGAGAGCCACCTCCAGCGCTTCTTAAGAGAGGTGA 132
QY 121 GGACTGAGACTTACAGAGCCCACTGCTGCCAGCTTTGTAAGCAAGAGAGCCCAAGAC 180
Db 133 GGACTGAGACTTACAGAGCCCACTGCTGCCAGCTTTGTAAGCAAGAGAGCCCAAGAC 192
QY 181 GCGCTTCCCGCGGCGGAGAGAGCCCACTGCTGCCAGCTTTGTAAGCAAGAGAGCCCAAGAC 240
Db 193 GCGCTTCCCGCGGCGGAGAGAGCCCACTGCTGCCAGCTTTGTAAGCAAGAGAGCCCAAGAC 252
QY 241 ACTGCGCGGCGGCGGAGAGAGCCCACTGCTGCCAGCTTTGTAAGCAAGAGAGCCCAAGAC 300
Db 253 ACTGCGCGGCGGCGGAGAGAGCCCACTGCTGCCAGCTTTGTAAGCAAGAGAGCCCAAGAC 312
QY 301 CGGATGCCCGAGGCTTCAAGCTTCCATGCTATTTATGTTTTTGGCTCA 360
Db 313 CGGATGCCCGAGGCTTCAAGCTTCCATGCTATTTATGTTTTTGGCTCA 372
QY 361 GCAGCAAGCTTCTGCTCATCTTCCGCGGAGTCCGTGCGCACTCG 405
Db 373 GCAGCAAGCTTCTGCTCATCTTCCGCGGAGTCCGTGCGCACTCG 417
RESULT 14
ADA53627/c
ID ADA53627 standard; cDNA; 2684 BP.
XX
XX ADA53627;
XX 20-NOV-2003 (first entry)
XX
XX Human coding sequence, SEQ ID 1195.
XX
XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX
XX Homo sapiens.
XX
XX EPI293569-A2.
XX
XX 19-MAR-2003.
XX
XX 21-MAR-2002; 2002EP-00006586.
PF

XX 14-SEP-2001; 2001JP-00328381.
PR 24-JAN-2002; 2002US-0350435P.
XX
XX (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Iasgai T, Sugiyama T, Otsuki T, Makamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;
XX WPI; 2003-395339/38.
DR P-PSDB; ADA55266.
XX
XX New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.
XX
XX Claim 1; SEQ ID NO 1195; 205pp; English.
PS
XX
XX The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
SQ Sequence 2684 BP; 599 A; 785 C; 684 G; 616 T; 0 U; 0 Other;
Query Match 27.2%; Score 392; DB 7; Length 2684;
Best Local Similarity 100.0%; Pred. No. 6-56-141;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1027 GGGCTCTGTGCTCTTCTTCGAGAGGCGGTGTGAGTCTTCAGTATGTTGCGCCAGC 1086
Db 2243 GGGCTCTGTGCTCTTCTTCGAGAGGCGGTGTGAGTCTTCAGTATGTTGCGCCAGC 2284
QY 1087 GCTCTTGGACCCCTTGTGACCAAGCGGCAAGAGCTGAGGCAAGAGAGAGGAGGCTCA 1146
Db 2283 GCTCTTGGACCCCTTGTGACCAAGCGGCAAGAGCTGAGGCAAGAGAGAGGAGGCTCA 2224
QY 1147 CCTCTTATCCTCGGCGACCCAGTGCACAGAGGCGCTCTCCAGACTTAAATGTATC 1206
Db 2223 CCTCTTATCCTCGGCGACCCAGTGCACAGAGGCGCTCTCCAGACTTAAATGTATC 2164
QY 1207 ACCACTTACCTGTGAGGAGGAGCCCAATCTGACTCTTCCCGCTTGGGACATGCGAG 1266
Db 2163 ACCACTTACCTGTGAGGAGGAGCCCAATCTGACTCTTCCCGCTTGGGACATGCGAG 2104
QY 1267 CCGGGAAGCAGTCCCGGCAAGGCTGTGGCCAGAGAGCTCCAGAGAGGCACTGACGCT 1326
Db 2103 CCGGGAAGCAGTCCCGGCAAGGCTGTGGCCAGAGAGCTCCAGAGAGGCACTGACGCT 2044
QY 1327 GCTGGCGGAGGCTTGGACATCCGAGGACACAGGAAAGTCTCTGGGGGATCTGTA 1386
Db 2043 GCTGGCGGAGGCTTGGACATCCGAGGACACAGGAAAGTCTCTGGGGGATCTGTA 1984
QY 1387 AATAACCTTTTCTTTTCTTTTAAATA 1418
Db 1983 AATAACCTTTTCTTTTCTTTTAAATA 1952
RESULT 15
AAS6563/c
ID AAS6563 standard; cDNA; 406 BP.
XX
XX AAS6563;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #2367.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
KW


```

XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US008631.
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
XX P-PSDB; AB902376.
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX Claim 1; SEQ ID NO 2367; 103bp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AS64197-AS64564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 406 BP; 83 A; 107 C; 138 G; 78 T; 0 U; 0 Other;
XX
QY 1 AAAGTAAACGGCTACAGACAGTGAAGAAATGTTTCGTCGCCGCTAGAAAAAAGTGTGCG 60
Db 405 AAAGTAAACGGCTACAGACAGTGAAGAAATGTTTCGTCGCCGCTAGAAAAAAGTGTGCG 346
QY 61 GTACCAACCCCAAGAGGCTTGAAGAGAGCCCACTCCAGCTTCTTAACGAGAGAGTGA 120
Db 345 GGAACCAACCCCAAGAGGCTTGAAGAGAGCCCACTCCAGCTTCTTAACGAGAGAGTGA 286
QY 121 GAAGTCAAGTCTTACCAAGCCCACTCGGTCCCAAGCTTGTAGCGAAAGAGAGCCCAAGAG 180
Db 285 GAAGTCAAGTCTTACCAAGCCCACTCGGTCCCAAGCTTGTAGCGAAAGAGAGCCCAAGAG 226
QY 181 GCGGTCTCCCGCGCTTCAAGAGAGCCCACTCGGTCCCAAGCTTGTAGCGAGG 240
Db 225 GCGGTCTCCCGCGCTTCAAGAGAGCCCACTCGGTCCCAAGCTTGTAGCGAGG 166
QY 241 ACTGAGCGGCGCTGACAGATGACCTGTGGAACGCGCTACTGCTTTTACCCCAAGGCC 300
Db 165 ACTGAGCGGCGCTGACAGATGACCTGTGGAACGCGCTACTGCTTTTACCCCAAGGCC 106

```

```

QY 301 CGGATGCCGAGGCTTCAAGCTTCACTGCTCATGTTATTTAGTGTGCTCTGA 360
Db 105 CGGATGCCGAGGCTTCAAGCTTCACTGCTCATGTTATTTAGTGTGCTCTGA 46
QY 361 GCAGCAAGCTTCTGCTCATCTTGGCGGGAGTCCGTGGCACTCG 405
Db 45 GCAGCAAGCTTCTGCTCATCTTGGCGGGAGTCCGTGGCACTCG 1

```

Search completed: February 23, 2004, 23:49:19
 CDD Time : 586.643 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 23:23:40 ; Search time 130.797 Seconds
(without alignments)
6113.919 Million cell updates/sec

Title: US-09-936-456-1
Perfect score: 1441
Sequence: 1 aaagtaacgctacagacag.....aaaaaaaaaaaaaaaa 1441

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_5/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_5/ptodata/2/ina/5A_COMB.seq: *
4: /cgn2_5/ptodata/2/ina/5B_COMB.seq: *
5: /cgn2_5/ptodata/2/ina/PTUS_COMB.seq: *
6: /cgn2_5/ptodata/2/ina/backfile1.seq: *

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	2.5	2255	3 US-08-871-572B-3	Sequence 3, Appli
2	35	2.4	1639	3 US-09-362-473-5	Sequence 5, Appli
3	34	2.4	144	1 US-08-702-344-26	Sequence 26, Appli
4	34	2.4	348	4 US-09-621-976-13740	Sequence 13740, A
5	34	2.4	388	4 US-09-621-976-18573	Sequence 18573, A
6	34	2.4	396	4 US-08-640-173-10	Sequence 10, Appli
7	34	2.4	396	4 US-08-713-550-10	Sequence 10, Appli
8	34	2.4	397	4 US-09-621-976-18571	Sequence 18571, A
9	34	2.4	769	4 US-08-567-882-5	Sequence 5, Appli
10	34	2.4	1191	3 US-09-282-305-13	Sequence 13, Appli
11	34	2.4	1191	4 US-09-883-720-13	Sequence 13, Appli
12	34	2.4	1283	1 US-08-174-467-19	Sequence 19, Appli
13	34	2.4	1283	2 US-08-452-071-19	Sequence 19, Appli
14	34	2.4	1518	3 US-09-614-912-191	Sequence 191, App
15	34	2.4	1582	3 US-08-545-1968-10	Sequence 10, Appli
16	34	2.4	1582	3 US-08-545-1968-12	Sequence 12, Appli
17	34	2.4	2043	4 US-09-614-912-171	Sequence 171, Appli
18	34	2.4	2103	4 US-09-489-847-40	Sequence 40, Appli
19	34	2.4	2218	2 US-08-985-090-4	Sequence 4, Appli
20	34	2.4	2218	3 US-09-165-543-31	Sequence 31, Appli
21	34	2.4	2780	4 US-09-489-847-47	Sequence 47, Appli
22	34	2.4	3244	3 US-08-165-543-4	Sequence 4, Appli
23	34	2.4	3437	3 US-08-860-339-17	Sequence 17, Appli
24	34	2.4	3437	3 US-09-573-629-17	Sequence 17, Appli
25	34	2.4	3581	2 US-08-738-349-1	Sequence 1, Appli
26	34	2.4	4494	4 US-09-620-312D-861	Sequence 861, App
27	33	2.3	229	4 US-09-621-976-10914	Sequence 10914, A

28	33	2.3	273	4 US-09-621-976-11078	Sequence 11078, A
29	33	2.3	375	3 US-08-946-026-23	Sequence 23, Appli
30	33	2.3	724	3 US-09-020-956-38	Sequence 38, Appli
31	33	2.3	724	4 US-09-030-607-38	Sequence 38, Appli
32	33	2.3	724	4 US-09-439-313-38	Sequence 38, Appli
33	33	2.3	724	4 US-09-352-616A-38	Sequence 38, Appli
34	33	2.3	724	4 US-09-232-149A-38	Sequence 38, Appli
35	33	2.3	724	4 US-09-159-812-38	Sequence 38, Appli
36	33	2.3	724	4 US-09-635-215-38	Sequence 38, Appli
37	33	2.3	724	4 US-09-685-166A-38	Sequence 38, Appli
38	33	2.3	724	4 US-09-115-453-38	Sequence 38, Appli
39	33	2.3	724	4 US-09-688-483-38	Sequence 38, Appli
40	33	2.3	911	2 US-08-924-759-9	Sequence 9, Appli
41	33	2.3	911	3 US-09-248-335-9	Sequence 9, Appli
42	33	2.3	1052	3 US-09-489-847-23	Sequence 23, Appli
43	33	2.3	1223	3 US-09-154-874-4	Sequence 4, Appli
44	33	2.3	1223	4 US-08-931-668-4	Sequence 4, Appli
45	33	2.3	1223	4 US-09-468-175-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-871-572B-3
Sequence 3, Application US/08871572B
Patent No. 6287853
GENERAL INFORMATION:
APPLICANT: Pestka, Sidney
APPLICANT: Korenko, Serguei
APPLICANT: Sob, Jaemog
APPLICANT: Domellay, Robert
APPLICANT: Maritano, Thomas
APPLICANT: Cook, Jeffrey
APPLICANT: Emmanuel, Stuart
APPLICANT: Schwartz, Barbara
TITLE OF INVENTION: Accessory Factor for Interferon Gamma
TITLE OF INVENTION: and Its Receptor
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Richard R. Muccino
STREET: 758 Springfield Avenue
CITY: Summit
STATE: New Jersey
COUNTRY: USA
ZIP: 07901
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/871,572B
FILING DATE: 9-JUNE-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Muccino, Richard R.
REGISTRATION NUMBER: 32,538
REFERENCE/DOCKET NUMBER: UMD1-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 273-4988
TELEFAX: (908) 273-4679
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2255 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-871-572B-3

Query Match 2.5%; Score 36; DB 3; Length 2255;
 Best Local Similarity 100.0%; Pred. No. 2.5e-06;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1406 TGTGTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1441
 DB 2209 TGTGTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2244

RESULT 2
 US-09-362-473-5
 ; Sequence 5, Application US/09362473
 ; Patent No. 6218169
 ; GENERAL INFORMATION:
 ; APPLICANT: Cahoon, Edgar B.
 ; APPLICANT: Cahoon, Rebecca E.
 ; APPLICANT: Falco, S. Carl
 ; APPLICANT: Morgan, Michele
 ; APPLICANT: Rafalski, J. Antoni
 ; APPLICANT: Hiltz, William D.
 ; APPLICANT: Kinney, Anthony J.
 ; TITLE OF INVENTION: Aromatic Amino Acid Catabolism Enzymes
 ; FILE REFERENCE: BR-1197
 ; CURRENT APPLICATION NUMBER: US/09/362,473
 ; EARLIER FILING DATE: 1999-07-28
 ; EARLIER APPLICATION NUMBER: 60/094,783
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 5
 ; LENGTH: 1639
 ; TYPE: DNA
 ; ORGANISM: Glycine max
 ; US-09-362-473-5

Query Match 2.4%; Score 35; DB 3; Length 1639;
 Best Local Similarity 100.0%; Pred. No. 7e-06;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1407 GTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1441
 DB 1591 GTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1625

RESULT 3
 US-08-702-344-26
 ; Sequence 26, Application US/08702344
 ; Patent No. 5723315
 ; GENERAL INFORMATION:
 ; APPLICANT: Jacobs, Kenneth
 ; APPLICANT: McCoy, John
 ; APPLICANT: Lavallee, Edward
 ; APPLICANT: Racine, Lisa
 ; APPLICANT: Weisberg, David
 ; APPLICANT: Treacy, Maurice
 ; APPLICANT: Spaulding, Vikki
 ; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
 ; TITLE OF INVENTION: ENCODING THEM
 ; NUMBER OF SEQUENCES: 37
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc.
 ; STREET: 87 CambridgePark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/702,344

;; FILING DATE:
 ;; CLASSIFICATION: 536
 ;; ATTORNEY/AGENT INFORMATION:
 ;; NAME: Brown, Scott A. 32,724
 ;; REGISTRATION NUMBER: 32,724
 ;; TELECOMMUNICATION INFORMATION:
 ;; TELEPHONE: (617) 498-8224
 ;; TELEFAX: (617) 876-5851
 ;; INFORMATION FOR SEQ ID NO: 26:
 ;; SEQUENCE CHARACTERISTICS:
 ;; LENGTH: 144 base pairs
 ;; TYPE: nucleic acid
 ;; STRANDEDNESS: double
 ;; TOPOLOGY: linear
 ;; MOLECULE TYPE: CDNA
 ; US-08-702-344-26

Query Match 2.4%; Score 34; DB 1; Length 144;
 Best Local Similarity 100.0%; Pred. No. 2.2e-05;
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1408 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1441
 DB 17 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 50

RESULT 4
 US-09-621-976-13740
 ; Sequence 13740, Application US/09621976
 ; Patent No. 6639063
 ; GENERAL INFORMATION:
 ; APPLICANT: Dumas Milne Edwards, J.B.
 ; APPLICANT: Uober, S.
 ; APPLICANT: Giordano, J.Y.
 ; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
 ; FILE REFERENCE: GENSET.054PR2
 ; CURRENT APPLICATION NUMBER: US/09/621,976
 ; CURRENT FILING DATE: 2000-07-21
 ; NUMBER OF SEQ ID NOS: 19335
 ; SOFTWARE: Patent.pm
 ; SEQ ID NO 13740
 ; LENGTH: 348
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-621-976-13740

Query Match 2.4%; Score 34; DB 4; Length 348;
 Best Local Similarity 100.0%; Pred. No. 2.1e-05;
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1408 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1441
 DB 314 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 347

RESULT 5
 US-09-621-976-18573
 ; Sequence 18573, Application US/09621976
 ; Patent No. 6639063
 ; GENERAL INFORMATION:
 ; APPLICANT: Dumas Milne Edwards, J.B.
 ; APPLICANT: Uober, S.
 ; APPLICANT: Giordano, J.Y.
 ; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
 ; FILE REFERENCE: GENSET.054PR2
 ; CURRENT APPLICATION NUMBER: US/09/621,976
 ; CURRENT FILING DATE: 2000-07-21
 ; NUMBER OF SEQ ID NOS: 19335
 ; SOFTWARE: Patent.pm
 ; SEQ ID NO 18573
 ; LENGTH: 368
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens

US-09-621-976-18573

Query Match 2.4%; Score 34; DB 4; Length 388;
Best Local Similarity 100.0%; Pred. No. 2.1e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1408 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1441
Db 343 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 376

RESULT 6

US-09-640-173-10/c
; Sequence 10, Application US/09640173
; Patent No. 6613515
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolck, John A.
; TITLE OF INVENTION: OVARIAN TUMOR SEQUENCES AND
; TITLE OF INVENTION: METHODS OF USE THEREFOR
; FILE REFERENCE: 210121.484C2
; CURRENT APPLICATION NUMBER: US/09/640.173
; CURRENT FILING DATE: 2000-08-15
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(396)
; OTHER INFORMATION: n = A,T,C or G
US-09-640-173-10

Query Match 2.4%; Score 34; DB 4; Length 396;
Best Local Similarity 100.0%; Pred. No. 2.1e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1408 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1441
Db 110 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 77

RESULT 7

US-09-713-550-10/c
; Sequence 10, Application US/09713550
; Patent No. 6617109
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolck, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.484C4
; CURRENT APPLICATION NUMBER: US/09/713.550
; CURRENT FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 205
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(396)
; OTHER INFORMATION: n = A,T,C or G
US-09-713-550-10

Query Match 2.4%; Score 34; DB 4; Length 396;
Best Local Similarity 100.0%; Pred. No. 2.1e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1408 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1441

Db 110 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 77

RESULT 8

US-09-621-976-18571
; Sequence 18571, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621.976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 18571
; LENGTH: 397
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-18571

Query Match 2.4%; Score 34; DB 4; Length 397;
Best Local Similarity 100.0%; Pred. No. 2.1e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1408 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1441
Db 343 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 376

RESULT 9

US-08-567-882-5
; Sequence 5, Application US/08567882
; Patent No. 6512103
; GENERAL INFORMATION:
; APPLICANT: Dairaghi, Daniel J.
; APPLICANT: Hara, Takahiko
; APPLICANT: Miyajima, Atsushi
; APPLICANT: Schall, Thomas J.
; APPLICANT: Wang, Wei
; APPLICANT: Yoshimura, Akihiko
; TITLE OF INVENTION: MAMMALIAN CHEMOKINE REAGENTS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/567.882
; FILING DATE: 08-DEC-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0506
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-852-9196
; TELEFAX: 415-496-1200
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 769 base pairs
; TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-567-882-5

Query Match 2.4%; Score 34; DB 4; Length 769;
Best Local Similarity 100.0%; Pred. No. 2e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1408 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1441
DB 523 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 556

RESULT 10
US-09-282-305-13
Sequence 13, Application US/09282305
Patent No. 6287843

GENERAL INFORMATION:
APPLICANT: Baldwin, Donald A.
APPLICANT: Briggs, Steven P.
APPLICANT: Crane, Virginia C.
TITLE OF INVENTION: Maize Histone Deacetylases And Their Uses
FILE REFERENCE: 5718-44,
CURRENT APPLICATION NUMBER: US/09/282,305
CURRENT FILING DATE: 1999-03-31
PRIOR APPLICATION NUMBER: 60/080,563
PRIOR FILING DATE: 1998-04-03
NUMBER OF SEQ ID NOS: 18
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 13
LENGTH: 1191
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: CDS
LOCATION: (63) ..(971)
US-09-282-305-13

Query Match 2.4%; Score 34; DB 3; Length 1191;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1408 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1441
DB 1158 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1191

RESULT 11
US-09-883-720-13
Sequence 13, Application US/09883720
Patent No. 6479629

GENERAL INFORMATION:
APPLICANT: Baldwin, Donald A.
APPLICANT: Briggs, Steven P.
APPLICANT: Crane, Virginia C.
TITLE OF INVENTION: Maize Histone Deacetylases And Their Uses
FILE REFERENCE: 5718-44,
CURRENT APPLICATION NUMBER: US/09/883,720
CURRENT FILING DATE: 2001-06-18
PRIOR APPLICATION NUMBER: 09/282,305
PRIOR FILING DATE: 1999-03-31
NUMBER OF SEQ ID NOS: 18
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 13
LENGTH: 1191
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: CDS
LOCATION: (63) ..(971)
US-09-883-720-13

Query Match 2.4%; Score 34; DB 4; Length 1191;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1408 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1441
DB 1158 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1191

RESULT 12
US-08-174-467-19
Sequence 19, Application US/08174467
Patent No. 5451514

GENERAL INFORMATION:
APPLICANT: BOUDET, ALAIN M.
APPLICANT: INZE, DIRK G.
APPLICANT: SCHUCH, WOLFGANG W.
TITLE OF INVENTION: MODIFICATION OF LIGNIN SYNTHESIS IN
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/174,467
FILING DATE: 28-DEC-1993
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/874,166
FILING DATE: 27-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 95563/PS36321/US

TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1283 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-174-467-19

Query Match 2.4%; Score 34; DB 1; Length 1283;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1408 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1441
DB 1244 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1277

RESULT 13
US-08-452-071-19
Sequence 19, Application US/08452071
Patent No. 6066760
GENERAL INFORMATION:
APPLICANT: BOUDET, ALAIN M.
APPLICANT: INZE, DIRK G.
APPLICANT: SCHUCH, WOLFGANG W.
TITLE OF INVENTION: MODIFICATION OF LIGNIN SYNTHESIS IN

```

; TITLE OF INVENTION: PLANTS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,071
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/874,166
; FILING DATE: 27-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 95563/PS36321/US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1283 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-08-452-071-19

Query Match 2.4%; Score 34; DB 3; Length 1283;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1408 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1441
Db 1244 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1277

RESULT 14
US-09-614-912-191
; Sequence 191, Application US/09614912
; Patent No. 6677502
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Rafalski, Antoni
; APPLICANT: Orozco, Buddy
; APPLICANT: Miao, Gou-Hau
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Lee, Jian Ming
; APPLICANT: Sakai, Hajime
; APPLICANT: Weng, Zude
; APPLICANT: Calmi, Perry G
; APPLICANT: Anderson, Shawn
; TITLE OF INVENTION: Plant Metabolism Genes
; FILE REFERENCE: B81378 US NA
; CURRENT APPLICATION NUMBER: US/09/614,912
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: 60/143,401
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/143,412
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/146,650
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 60/110,906
; PRIOR FILING DATE: 1999-12-15
```

```

; PRIOR APPLICATION NUMBER: 60/172,959
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/172,946
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 204
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 191
; LENGTH: 1518
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (114)
; NAME/KEY: unsure
; LOCATION: (123)
; NAME/KEY: unsure
; LOCATION: (118)
; US-09-614-912-191

Query Match 2.4%; Score 34; DB 4; Length 1518;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1408 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1441
Db 1434 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1467

RESULT 15
US-08-545-196B-10
; Sequence 10, Application US/08545196B
; Patent No. 6080577
; GENERAL INFORMATION:
; APPLICANT: MELKI, JUDITH
; APPLICANT: MUNNICH, ARNOLD
; TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE
; TITLE OF INVENTION: FOR SPINAL MUSCULAR ATROPHY
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,196B
; FILING DATE: 19-OCT-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FARACI, C. J.
; REGISTRATION NUMBER: 32,350
; REFERENCE/DOCKET NUMBER: 2121-110P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8050
; TELEFAX: (703) 205-8000
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1582 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-08-545-196B-10

Query Match 2.4%; Score 34; DB 3; Length 1582;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 1408 TTTTAAAAAAAAAAAAAAAAAAAAAAAA 1441
Db 1485 TTTTAAAAAAAAAAAAAAAAAAAAAAAA 1518

Search completed: February 24, 2004, 03:24:16
Job time : 132.797 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 23:36:30 ; Search time 552.173 Seconds
(without alignments)
9138.208 Million cell updates/sec

Title: US-09-936-456-1
Perfect score: 1441
Sequence: 1 aaagtaacgctacagacag.....aaaaaaaaaaaaaaaa 1441

Scoring table: OLIGO NUC
Gapop 60.0, Gapext 60.0

Searched: 2308684 seqs, 1750822206 residues

Word size: 0

Total number of hits satisfying chosen parameters: 4617368

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

Published Applications NA:*

- 1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq:*
- 3: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:*
- 7: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/prodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/prodata/1/pubpna/US09C_PUBCOMB.seq:*
- 13: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
- 14: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
- 15: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1413	98.1	1421	US-10-187-657-2
2	1404	97.4	1420	US-10-187-657-4
3	1285	89.2	1474	US-10-264-237-1097
4	392	27.2	2684	US-10-094-749-1195
5	313	21.7	5222	US-10-187-657-5
6	187	13.0	232	US-09-783-590-9492
7	136	9.4	346	US-10-187-657-3
8	119	8.3	450	US-10-187-657-7
9	107	7.4	439	US-10-027-632-91970
10	59	4.1	512	US-10-106-698-3277
11	56	3.9	506	US-10-187-657-6
12	38	2.6	1398	US-10-424-599-64755
13	38	2.6	2165	US-10-116-255-18
14	38	2.6	3095	US-10-424-599-75074
15	38	2.6	5642	US-10-311-455-842

16	2.6	10480	14	US-10-311-455-2174	Sequence 2174, Ap
17	2.6	554	12	US-10-424-599-41481	Sequence 4149, A
18	2.6	2464	12	US-10-424-599-8115	Sequence 8115, Ap
19	2.6	367378	14	US-10-312-841-1	Sequence 1, Appl
20	2.5	896	14	US-10-106-698-234	Sequence 234, Appl
21	2.4	290	15	US-10-242-5354-52594	Sequence 52594, A
22	2.4	302	12	US-10-424-599-48927	Sequence 48927, A
23	2.4	379	12	US-10-424-599-43341	Sequence 43341, A
24	2.4	379	12	US-10-424-599-133773	Sequence 133773, A
25	2.4	416	12	US-10-424-599-50281	Sequence 50281, A
26	2.4	449	12	US-10-424-599-1341	Sequence 1341, Ap
27	2.4	450	14	US-10-196-846-13053	Sequence 13053, A
28	2.4	450	15	US-10-131-827-6647	Sequence 8647, Ap
29	2.4	464	15	US-10-131-827-6179	Sequence 8179, Ap
30	2.4	481	12	US-10-424-599-120735	Sequence 120735, A
31	2.4	482	12	US-10-424-599-96207	Sequence 96207, A
32	2.4	484	12	US-10-424-599-17387	Sequence 17387, A
33	2.4	498	12	US-10-424-599-17873	Sequence 17873, A
34	2.4	509	12	US-10-424-599-107007	Sequence 107007, A
35	2.4	529	14	US-10-424-599-3130	Sequence 3130, Ap
36	2.4	621	12	US-10-424-599-10941	Sequence 10941, A
37	2.4	658	12	US-10-424-599-122157	Sequence 122157, A
38	2.4	1167	15	US-10-027-632-260991	Sequence 260991, A
39	2.4	1639	9	US-09-742-958-5	Sequence 5, Appl
40	2.4	1758	12	US-10-424-599-91053	Sequence 91053, A
41	2.4	50	9	US-09-783-590-6670	Sequence 6670, Ap
42	2.4	60	13	US-10-057-940-14	Sequence 14, Appl
43	2.4	64	13	US-10-057-940-1	Sequence 1, Appl
44	2.4	64	13	US-10-057-940-1	Sequence 1, Appl
45	2.4	64	13	US-10-057-940-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-10-187-657-2
; Sequence 2, Application US/10187657
; Publication No. US20030068311A1
; GENERAL INFORMATION:
; APPLICANT: Laeak, Amy K.W.
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Azimzai, Yaida
; TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
; FILE REFERENCE: PV-0009 CIP
; CURRENT APPLICATION NUMBER: US/10/187,657
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: PCT/US00/07817
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/139,565
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program
; SEQ ID NO 2
; LENGTH: 1421
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Inctye ID No. US20030068311A1 4901066CB1
US-10-187-657-2

Query Match 98.1%; Score 1413; DB 14; Length 1421;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	AAAGTAACGCTACAGAGCTGAGTAATGTTTCGTCGCGCTAGAAACTGTCG	60
DB	9	AAAGTAACGCTACAGAGCTGAGTAATGTTTCGTCGCGCTAGAAACTGTCG	68
QY	61	GTACCAACCCAGAGGCTTAGAGAGCACCACCTCCACGCTTCTTAACGAGAGTGCA	120
DB	69	GTACCAACCCAGAGGCTTAGAGAGCACCACCTCCACGCTTCTTAACGAGAGTGCA	128

QY 121 GGACTCAGACTTACACAGCCGACCTGGTCCAGCCTGTAGCAAGAAAGAGCCCAAGAC 180
 DB 129 GGACTCAGACTTACACAGCCGACCTGGTCCAGCCTGTAGCAAGAAAGAGCCCAAGAC 188
 QY 181 GGGCTCTCCCGGCTCCAGGACGCCAGCTTGGCTTGGCTGGCCGCTGGCTGAGC 240
 DB 189 GGGCTCTCCCGGCTCCAGGACGCCAGCTTGGCTTGGCTGGCCGCTGGCTGAGC 248
 QY 241 ACTCGGCGGGGCTGAGCACTGACCTGTGAGAAAGGCGATAGCTTTTACCCAGACC 300
 DB 249 ACTCGGCGGGGCTGAGCACTGACCTGTGAGAAAGGCGATAGCTTTTACCCAGACC 308
 QY 301 CGGCAATGCGGAGGCTTACAGCCTTCACTGCTCATGTTATTTAGTGTGCTCTA 360
 DB 309 CGGCAATGCGGAGGCTTACAGCCTTCACTGCTCATGTTATTTAGTGTGCTCTA 368
 QY 361 GCAGCAAGCTTCTGCTCATCTTGGCGGGGATCCGCGCACTGCGCTGGTTGGTTG 420
 DB 369 GCAGCAAGCTTCTGCTCATCTTGGCGGGGATCCGCGCACTGCGCTGGTTGGTTG 428
 QY 421 GTGAGAGTTCTTCTGAGTCTGTGATAGGCGAGAAATTTGGCTGCTGACTTCACTGCA 480
 DB 429 GTGAGAGTTCTTCTGAGTCTGTGATAGGCGAGAAATTTGGCTGCTGACTTCACTGCA 488
 QY 481 GAATGTTCTGAGTACAGTGAACACCAACATCTCTCAAAAGCCTTCAAGCGCAGCGGC 540
 DB 489 GAATGTTCTGAGTACAGTGAACACCAACATCTCTCAAAAGCCTTCAAGCGCAGCGGC 548
 QY 541 GTTACAGCCCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
 DB 549 GTTACAGCCCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 608
 QY 601 ACCCGAGTGCATGAGTGAACAGACCATTTGATCAACAGAGCACTTCACTGCTGCTG 660
 DB 609 ACCCGAGTGCATGAGTGAACAGACCATTTGATCAACAGAGCACTTCACTGCTGCTG 668
 QY 661 AAGAGAAATTAAGCCCGGAGATGAGCGGAAAGCACTGAGAAAGGGGCTCGGACCCATG 720
 DB 669 AAGAGAAATTAAGCCCGGAGATGAGCGGAAAGCACTGAGAAAGGGGCTCGGACCCATG 728
 QY 721 CTCTACTGCGGAGAAAGTTACACACGAGTAGGCTTGGCTGCTGCTGCTGCTGCTGCTG 780
 DB 729 CTCTACTGCGGAGAAAGTTACACACGAGTAGGCTTGGCTGCTGCTGCTGCTGCTGCTG 788
 QY 781 CTGGCGGAGCACTAGCTTGGCGCAAGCTATGAGTGGCTTCTGCTTCTGCTTCTCTCC 840
 DB 789 CTGGCGGAGCACTAGCTTGGCGCAAGCTATGAGTGGCTTCTGCTTCTGCTTCTCTCC 848
 QY 841 AAGTGTGCTTCCAGGCGGCGCTCTACGAGGCTGAGGCTGAGCACTGCTGAGCAACCGGA 900
 DB 849 AAGTGTGCTTCCAGGCGGCGCTCTACGAGGCTGAGGCTGAGCACTGCTGAGCAACCGGA 908
 QY 901 GCGTTGCGGCTCTTGGGAGTCTTGGCGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 960
 DB 909 GCGTTGCGGCTCTTGGGAGTCTTGGCGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 968
 QY 961 CTGGCGTAGGCTCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGG 1020
 DB 969 CTGGCGTAGGCTCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGG 1028
 QY 1021 GCAACCGGCGCTTGGCTTGGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGG 1080
 DB 1029 GCAACCGGCGCTTGGCTTGGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGG 1088
 QY 1081 CCGAGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGG 1140
 DB 1089 CCGAGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGG 1148
 QY 1141 GGGTACCTCTTATCTCGGCGAGCCCACTGCAAGAGAGCGGCTCTCCAGACTTTAAA 1200
 DB 1149 GGGTACCTCTTATCTCGGCGAGCCCACTGCAAGAGAGCGGCTCTCCAGACTTTAAA 1208

QY 1201 TGATACCACTAAGTGTGAGGGGAGCCCAATCTGAGCTCTTCCCGCCTTGGAGCAT 1260
 DB 1209 TGATACCACTAAGTGTGAGGGGAGCCCAATCTGAGCTCTTCCCGCCTTGGAGCAT 1268
 QY 1261 CGCAGCGCGGAGAGAGTGTCCCGCAGGCTTGGCGCAGAGAGCTTCAAGAAAGGAGACTG 1320
 DB 1269 CGCAGCGCGGAGAGAGTGTCCCGCAGGCTTGGCGCAGAGAGCTTCAAGAAAGGAGACTG 1328
 QY 1321 AGCGTGTGCGGAGAGGCTTGGAGATCCGAGAGGACAGAGGAAAGTCTCTGGAGGCA 1380
 DB 1329 AGCGTGTGCGGAGAGGCTTGGAGATCCGAGAGGACAGAGGAAAGTCTCTGGAGGCA 1388
 QY 1381 TCTGTAAATAAATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1413
 DB 1389 TCTGTAAATAAATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1421
 RESULT 2
 US-10-187-657-4
 ; Sequence 4, Application US/10187657
 ; Publication No. US2003006831A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Iasek, Amy K.W.
 ; APPLICANT: Baughn, Mariah R.
 ; APPLICANT: Azimzal, Yalda
 ; TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
 ; FILE REFERENCE: PV-0009 CIP
 ; CURRENT APPLICATION NUMBER: US/10/187,657
 ; PRIOR FILING DATE: 2002-07-01
 ; PRIOR APPLICATION NUMBER: PCT/US00/07817
 ; PRIOR FILING DATE: 2000-03-22
 ; PRIOR APPLICATION NUMBER: 60/139,565
 ; PRIOR FILING DATE: 1999-06-16
 ; NUMBER OF SEQ ID NOS: 9
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 4
 ; LENGTH: 1420
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc.feature
 ; OTHER INFORMATION: Incyte ID No. US2003006831A1 322161CA2
 US-10-187-657-4
 Query Match 97.4%; Score 1404; DB 14; Length 1420;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAAGTAAAGGCTTACAGAGAGTGAAGAAATGTTGGCTGCGCGGCTTGAAGAAATCTGTG 60
 DB 9 AAAGTAAAGGCTTACAGAGAGTGAAGAAATGTTGGCTGCGCGGCTTGAAGAAATCTGTG 68
 QY 61 GTAACCAACCCAGAGAGGTTGAGAGAGCCCACTTCCAGGCTTCTTAAACGAGAGGTGA 120
 DB 69 GTAACCAACCCAGAGAGGTTGAGAGAGCCCACTTCCAGGCTTCTTAAACGAGAGGTGA 128
 QY 121 GGACTAGACTTACACAGCCCACTGAGTCCAGAGCTTGTAGCAAAAGAGAGCCAGAGAC 180
 DB 129 GGACTAGACTTACACAGCCCACTGAGTCCAGAGCTTGTAGCAAAAGAGAGCCAGAGAC 188
 QY 181 GGGCTTCCCGGCTCCAGGAGCCCACTTGGCTTGGCTGCTGCGCGCTTGGCTGAGC 240
 DB 189 GGGCTTCCCGGCTCCAGGAGCCCACTTGGCTTGGCTGCTGCGCGCTTGGCTGAGC 248
 QY 241 ACTCGGCGGGGCTGAGCACTGACCTGTGAGAAAGGCGATAGCTTTTACCCAGACC 300
 DB 249 ACTCGGCGGGGCTGAGCACTGACCTGTGAGAAAGGCGATAGCTTTTACCCAGACC 308
 QY 301 CGGCAATGCGGAGGCTTACAGCCTTCACTGCTCATGTTATTTAGTGTGCTCTA 360
 DB 309 CGGCAATGCGGAGGCTTACAGCCTTCACTGCTCATGTTATTTAGTGTGCTCTA 368
 QY 361 GCAGCAAGCTTCTGCTCATCTTGGCGGGGATCCGCGCACTGCGCTGGTTGGTTG 420

Db 369 GCAGCAAGCTTCTCCGCTCACTCTGCGGGGATCCGTGGCACTCGCGCTGGTTGGTTG 428
QY 421 GTGAGATTCTTCTCACTCTGTTCTATAGCGGCAAAATTGAGCTGTGCACTTCAATGCA 480
Db 429 GTGAGATTCTTCTCACTCTGTTCTATAGCGGCAAAATTGAGCTGTGCACTTCAATGCA 488
QY 481 GAATGTTCTGGGTATAGTAAACCAACACATCTCTCAAAAGCTTCAAGCGGAGCGG 540
Db 489 GAATGTTCTGGGTATAGTAAACCAACACATCTCTCAAAAGCTTCAAGCGGAGCGG 548
QY 541 GTTACAGCCCGTGTGGTGTCTCTGTGGGCTTGAAGGCAATTATATTAACCTCAAGG 600
Db 549 GTTACAGCCCGTGTGGTGTCTCTGTGGGCTTGAAGGCAATTATATTAACCTCAAGG 608
QY 601 ACCCAAGTGCATGAGTGAACAGAACCAATTGACTTACAAAGAGAGTTCACTGGGCTG 660
Db 609 ACCCAAGTGCATGAGTGAACAGAACCAATTGACTTACAAAGAGAGTTCACTGGGCTG 668
QY 661 AAAGAAATTACGCGCGGAGTACGCGAACGCACTGGAGAGGGGCTGCGGACCCAGTG 720
Db 669 AAAGAAATTACGCGCGGAGTACGCGAACGCACTGGAGAGGGGCTGCGGACCCAGTG 728
QY 721 CTCTACCTGCGGAGAAATTACACCGAGTACCCCTTGGGCTGTATACCAAGTATCCAC 780
Db 729 CTCTACCTGCGGAGAAATTACACCGAGTACCCCTTGGGCTGTATACCAAGTATCCAC 788
QY 781 CTGGCGGAGACATACGCGCTGGGAGCAAGCTATAGGTTGGGCTTCTGCTTGGGCTCCCTCC 840
Db 789 CTGGCGGAGACATACGCGCTGGGAGCAAGCTATAGGTTGGGCTTCTGCTTGGGCTCCCTCC 848
QY 841 AACGTGTCTCTCAAGCGCGCGCGCTCTACGGAAGGCTTGGCACTGTGACCAACGGA 900
Db 849 AACGTGTCTCTCAAGCGCGCGCGCTCTACGGAAGGCTTGGCACTGTGACCAACGGA 908
QY 901 GCGTTCGCGCTCTTGGGCTTGTGGCTTGTGGCTTGTGGCTTGTGGCTTGTGGCTTGTGG 960
Db 909 GCGTTCGCGCTCTTGGGCTTGTGGCTTGTGGCTTGTGGCTTGTGGCTTGTGGCTTGTGG 968
QY 961 CTCCGCTTACGCTCTCCGCGCTCAACATCAATAGAGCGCGCTTCTGGTCAAGCTG 1020
Db 969 CTCCGCTTACGCTCTCCGCGCTCAACATCAATAGAGCGCGCTTCTGGTCAAGCTG 1028
QY 1021 GCAACCGCGCTCTGTGTGCTTCTCTGGAAGGCGCGCTGTGATCTCCAGTATGTTCCG 1080
Db 1029 GCAACCGCGCTCTGTGTGCTTCTCTGGAAGGCGCGCTGTGATCTCCAGTATGTTCCG 1088
QY 1081 CCCAGCGCTCTTGGCACTCTTCTGGAACCAAGCGCGCAAGACTGACGAGAGAGG 1140
Db 1089 CCCAGCGCTCTTGGCACTCTTCTGGAACCAAGCGCGCAAGACTGACGAGAGAGG 1148
QY 1141 GCGTCACTCTTATCTCGCGGAGCCCACTGCAAGAGAGCGCTTCTCCAGACTTAA 1200
Db 1149 GCGTCACTCTTATCTCGCGGAGCCCACTGCAAGAGAGCGCTTCTCCAGACTTAA 1208
QY 1201 TGTATCACCACTAATCTGTGAAGGAGGAGCCCAATGTGACTCTTCCCGCTTGGGCAAT 1260
Db 1209 TGTATCACCACTAATCTGTGAAGGAGGAGCCCAATGTGACTCTTCCCGCTTGGGCAAT 1268
QY 1261 CGCAGCGCGGAGAGAGTACCGCGCAAGGCTTGGGCAAGAGAGTCTCCAGAGAGGCACTG 1320
Db 1269 CGCAGCGCGGAGAGAGTACCGCGCAAGGCTTGGGCAAGAGAGTCTCCAGAGAGGCACTG 1328
QY 1321 AACGCTGTGGGCGAGGCTTGGCAATCCGCAAGGAGCGCAAGAGAGAGTCTCCGAGGCGA 1380
Db 1329 AACGCTGTGGGCGAGGCTTGGCAATCCGCAAGGAGCGCAAGAGAGAGTCTCCGAGGCGA 1388
QY 1381 TCTGTAAATTAACCTTTTCTT 1404
Db 1389 TCTGTAAATTAACCTTTTCTT 1412

RESULT 3

US-10-264-237-1097
; Sequence 1097, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
; APPLICANT: Birst et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P4131P1
; CURRENT APPLICATION NUMBER: US/10/264,237
; PRIOR FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: Patentin Ver. 3.1
; SEQ ID NO 1097
; LENGTH: 1474
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-264-237-1097
Query Match 89.2%; Score 1285; DB 15; Length 1474;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1435; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAGTAAAGGCTTACAGACAGTGAAGAAATGTTGCTGCGCGCTTGAAGAAAATCTGTGCG 60
Db 27 AAAGTAAAGGCTTACAGACAGTGAAGAAATGTTGCTGCGCGCTTGAAGAAAATCTGTGCG 86
QY 61 GTACCAACCCCAAGAGGTTGAGAGGAGCCCACTCCAGCTTCTTAAAGAGAGGTGA 120
Db 87 GTACCAACCCCAAGAGGTTGAGAGGAGCCCACTCCAGCTTCTTAAAGAGAGGTGA 146
QY 121 GAACTCACTTACAGACAGCCCACTCGGTCCAGCTTGTACCAAGAGAGGCAAGAC 180
Db 147 GAACTCACTTACAGACAGCCCACTCGGTCCAGCTTGTACCAAGAGAGGCAAGAC 206
QY 181 GCGCTCTCCGCGCTTCCAGAGGAGCCCACTGTGCTGCTGCGCGCGCTGAGGAGC 240
Db 207 GCGCTCTCCGCGCTTCCAGAGGAGCCCACTGTGCTGCTGCGCGCGCTGAGGAGC 266
QY 241 ACTGCGCGCGCTTCCAGAGGAGCCCACTGTGCTGCTGCGCGCGCTGAGGAGC 300
Db 267 ACTGCGCGCGCTTCCAGAGGAGCCCACTGTGCTGCTGCGCGCGCTGAGGAGC 326
QY 301 CGGATATCCGAGGCTTCAAGCTTCACTGCTCATCTGATGTTATGTTGTTGCTCTA 360
Db 327 CGGATATCCGAGGCTTCAAGCTTCACTGCTCATCTGATGTTATGTTGTTGCTCTA 386
QY 361 GCAGCAAGCTTCCGCTCATCTTCCGCGGAGTCCGTGCGCACTCGCGCTGTTTGGTTG 420
Db 387 GCAGCAAGCTTCCGCTCATCTTCCGCGGAGTCCGTGCGCACTCGCGCTGTTTGGTTG 446
QY 421 GTGAGAGTCTTCTCACTCTGTTCTATGAGCGGCAAAATTGAGCTGTGCACTTCAATGCA 480
Db 447 GTGAGAGTCTTCTCACTCTGTTCTATGAGCGGCAAAATTGAGCTGTGCACTTCAATGCA 506
QY 481 GAATGTTCTGGGTATAGTAAACCAACACATCTCTCAAAAGCTTCAAGCGGAGCGG 540
Db 507 GAATGTTCTGGGTATAGTAAACCAACACATCTCTCAAAAGCTTCAAGCGGAGCGG 566
QY 541 GTTACAGCCCGTGTGGTGTCTCTGTGGGCTTGAAGGCAATTATTAACCTCAAGG 600
Db 567 GTTACAGCCCGTGTGGTGTCTCTGTGGGCTTGAAGGCAATTATTAACCTCAAGG 626
QY 601 ACCCAAGTGCATGAGTGAACAGAACCAATTGACTTACAAAGAGAGTTCACTGGGCTG 660
Db 627 ACCCAAGTGCATGAGTGAACAGAACCAATTGACTTACAAAGAGAGTTCACTGGGCTG 686
QY 661 AAAGAAATTACGCGCGGAGTACGCGAACGCACTGGAGAGGGGCTGCGGACCCAGTG 720
Db 687 AAAGAAATTACGCGCGGAGTACGCGAACGCACTGGAGAGGGGCTGCGGACCCAGTG 746

QY 721 CTCTACCTGCGGAGAAAGTTCAACAGAGTACCCCTTGGGCTGTACACACAGTACAC 780
DB 747 CTCTACCTGCGGAGAAAGTTCAACAGAGTACCCCTTGGGCTGTACACACAGTACAC 806
QY 781 CTGCGGAGACACTAGCCCTCGGCAAGTATGAGTGGGCTTCTGCTTCTGCTCTCTCC 840
DB 807 CTGCGGAGACACTAGCCCTCGGCAAGTATGAGTGGGCTTCTGCTTCTGCTCTCTCC 866
QY 841 AACGTGCTCTCTCCACGCGGCGCCGCTTACAGAGGCTTGGACATGCTTACACCGGA 900
DB 867 AACGTGCTCTCTCCACGCGGCGCCGCTTACAGAGGCTTGGACATGCTTACACCGGA 926
QY 901 GCCTTGCGCTCTTCTGCGGCTCTTCTGCGCTTCTGCGCTTCTGCGCTTCTGCGCG 960
DB 927 GCCTTGCGCTCTTCTGCGGCTCTTCTGCGCTTCTGCGCTTCTGCGCTTCTGCGCG 986
QY 961 CTGCGCTTGGCTCTCTCGGCGCTTCACTACAGTACGCGCGCTTCTGAGTACGCTG 1020
DB 987 CTGCGCTTGGCTCTCTCGGCGCTTCACTACAGTACGCGCGCTTCTGAGTACGCTG 1046
QY 1021 GCAACCGGCGCTCTCTGCTCTCTCTGCGAGGCGCTGAGTCTCAATATGTTCTG 1080
DB 1047 GCAACCGGCGCTCTCTGCTCTCTCTGCGAGGCGCTGAGTCTCAATATGTTCTG 1106
QY 1081 CCCAGCGCTCTTGGCACTCTTCTGCAACCAAGCGCAAGACTGAGCCAGAGAGAGG 1140
DB 1107 CCCAGCGCTCTTGGCACTCTTCTGCAACCAAGCGCAAGACTGAGCCAGAGAGAGG 1166
QY 1141 GGCTCACCTTATCTCTGCGAGACCACTGCAAGAGGCGCTCTCCAGACTTAA 1200
DB 1167 GGCTCACCTTATCTCTGCGAGACCACTGCAAGAGGCGCTCTCCAGACTTAA 1226
QY 1201 TGTATCACCACTTACCTGTGAGGAGGAGCCCAATCTGCACTCTTCTCCGCTTGGACAT 1260
DB 1227 TGTATCACCACTTACCTGTGAGGAGGAGCCCAATCTGCACTCTTCTCCGCTTGGACAT 1286
QY 1261 CGAGGCGGAGAGAGAGTCCGCGAGCGCTGCGAGAGGCTCGAGAGAGGAGCACTG 1320
DB 1287 CGAGGCGGAGAGAGAGTCCGCGAGCGCTGCGAGAGGCTCGAGAGAGGAGCACTG 1346
QY 1321 AGCGTGTGCGCGGAGGCGCTCGACATCCGAGGAGCAACAGAGAAAGTCTCTGAGGAG 1380
DB 1347 AGCGTGTGCGCGGAGGCGCTCGACATCCGAGGAGCAACAGAGAAAGTCTCTGAGGAG 1406
QY 1381 TCTGTAAATPAACTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1438
DB 1407 TCTGTAAATPAACTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1464

RESULT 4.
US-10-094-749-1195/c
Sequence 1195, Application US/10094749
Publication No. US20030219741A1

GENERAL INFORMATION:

APPLICANT: ISOGAI, TAKAO
APPLICANT: SUGIYAMA, TOMOYASU
APPLICANT: OTSUKI, TETSUJI
APPLICANT: WAKAMATSU, AI
APPLICANT: SATO, HIROYUKI
APPLICANT: ISHII, SHIZUKO
APPLICANT: YAMAMOTO, JUN-ICHI
APPLICANT: ISONO, YUUKO
APPLICANT: HIO, YURI
APPLICANT: OTSUKA, KAORU
APPLICANT: NAGAI, KEIICHI
APPLICANT: IRIE, RYOTARO
APPLICANT: TAMECHIKA, ICHIRO
APPLICANT: SEKI, NAOHICO
APPLICANT: YOSHIKAWA, TSUTOMU
APPLICANT: OTSUKA, MOTOTUKI
APPLICANT: NAGAHARI, KENJI
APPLICANT: MASUHO, YASUHIKO
TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA

FILE REFERENCE: 084335/0160
CURRENT APPLICATION NUMBER: US/10/094,749
CURRENT FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 60/350,435
PRIOR FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: JP 2001-328381
PRIOR FILING DATE: 2001-09-14
NUMBER OF SEQ ID NOS: 3381
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 1195
LENGTH: 2684
TYPE: DNA
ORGANISM: Homo sapiens
US-10-094-749-1195

Query Match 27.2%; Score 392; DB 15; Length 2684;
Best Local Similarity 100.0%; Pred. No. 3e-181;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1027 GGCGTCTGTGCTCTTCTCTGAGAGGCGCTGTGAGTCTCAGATATGTTGGCCGAC 1086
DB 2343 GGCGTCTGTGCTCTTCTCTGAGAGGCGCTGTGAGTCTCAGATATGTTGGCCGAC 2284
QY 1087 GCTTTGGCACTCTTCTGAGCAAGAGGCGCAAGAGTCTGAGCGAGAGAGAGGAGCTCA 1146
DB 2283 GCTTTGGCACTCTTCTGAGCAAGAGGCGCAAGAGTCTGAGCGAGAGAGAGGAGCTCA 2224
QY 1147 CCTTTTATCTCTGAGAGCCCACTGCAAGAGGCGCTCTCCAGACTTAAATGATATC 1206
DB 2223 CCTTTTATCTCTGAGAGCCCACTGCAAGAGGCGCTCTCCAGACTTAAATGATATC 2164
QY 1207 ACCACTTACCTGTGAGGAGGAGCCCAATCTGAGTCTTCTCCGCTTGGAGCAATCCGAGG 1266
DB 2163 ACCACTTACCTGTGAGGAGGAGCCCAATCTGAGTCTTCTCCGCTTGGAGCAATCCGAGG 2104
QY 1267 CGGAGAGAGTGGCGCGCGAGGCGCTGAGGAGAGTCTCAGAGAGGAGGAGCACTGAGCGCT 1326
DB 2103 CGGAGAGAGTGGCGCGCGAGGCGCTGAGGAGAGTCTCAGAGAGGAGGAGCACTGAGCGCT 2044
QY 1327 GCTGGCGGAGAGGCTCTGAGCAATCCGAGGAGCAACAGGAGAAAGTCTCTTGGGCGCATCTGTA 1386
DB 2043 GCTGGCGGAGAGGCTCTGAGCAATCCGAGGAGCAACAGGAGAAAGTCTCTTGGGCGCATCTGTA 1984
QY 1387 AATAAACCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1418
DB 1983 AATAAACCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1952

RESULT 5
US-10-187-657-5
Sequence 5, Application US/10187657
Publication No. US2003006831A1

GENERAL INFORMATION:

APPLICANT: Lasek, Amy K.W.
APPLICANT: Baughn, Marian R.
APPLICANT: Azimzal, Yalda
TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
FILE REFERENCE: PV-0009 CIP
CURRENT APPLICATION NUMBER: US/10/187,657
CURRENT FILING DATE: 2002-07-01
PRIOR APPLICATION NUMBER: PCT/US00/07817
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60/139,565
PRIOR FILING DATE: 1999-06-16
SOFTWARE: PERL Program
SEQ ID NO: 5
LENGTH: 522
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. US2003006831A1 322166186

FEATURE:
NAME/KEY: unsure
LOCATION: 455, 480, 483
OTHER INFORMATION: a, t, c, g, or other
US-10-187-657-5

Query Match 21.7%; Score 313; DB 14; Length 522;
Best Local Similarity 100.0%; Pred. No. 1.4e-142; Indels 0; Gaps 0;
Matches 313; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAGTAAAGGCTTACAGACAGTGAAGAAATGTTGCTGCGCGGCTTAAAGAAAATCTGTG 60
DB 9 AAAGTAAAGGCTTACAGACAGTGAAGAAATGTTGCTGCGCGGCTTAAAGAAAATCTGTG 68
QY 61 GTACCAACCCGAGGCGTTGAGAGAGCCCACTCCAGCTTCTTAAAGAGAGGTGCA 120
DB 69 GTACCAACCCGAGGCGTTGAGAGAGCCCACTCCAGCTTCTTAAAGAGAGGTGCA 128
QY 121 GGAAGTAAAGGCTTACAGACAGTGAAGAAATGTTGCTGCGCGGCTTAAAGAAAATCTGTG 180
DB 129 GGAAGTAAAGGCTTACAGACAGTGAAGAAATGTTGCTGCGCGGCTTAAAGAAAATCTGTG 188
QY 181 GCGGCTTCCCGGCTTCAAGAGAGCCCACTCCAGCTTCTTAAAGAGAGGTGCA 240
DB 189 GCGGCTTCCCGGCTTCAAGAGAGCCCACTCCAGCTTCTTAAAGAGAGGTGCA 248
QY 241 ACTGAGGCGGCTTCAAGAGAGCCCACTCCAGCTTCTTAAAGAGAGGTGCA 300
DB 249 ACTGAGGCGGCTTCAAGAGAGCCCACTCCAGCTTCTTAAAGAGAGGTGCA 308
QY 301 CGGATGCGCGAG 313
DB 309 CGGATGCGCGAG 321

RESULT 6
US-09-783-590-9492
Sequence 9492, Application US/09783590
Patent No. US20020110850A1
GENERAL INFORMATION:
APPLICANT: Dillon, Patrick J.
APPLICANT: Haseltine, William A.
APPLICANT: Li, Haodong
APPLICANT: Rosen, Craig A.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
FILE REFERENCE: PO-16,2C1
CURRENT APPLICATION NUMBER: US/09/783,590
PRIORITY FILING DATE: 2000-02-15
PRIORITY APPLICATION NUMBER: 08/420,856
PRIORITY FILING DATE: 1995-04-12
PRIORITY APPLICATION NUMBER: 08/346,731
PRIORITY FILING DATE: 1994-11-21
NUMBER OF SEQ ID NOS: 12485
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 9492
LENGTH: 232
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (202)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (206)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (229)
OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-9492

Query Match 13.0%; Score 187; DB 9; Length 232;
Best Local Similarity 100.0%; Pred. No. 5.6e-81;

Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAGTAAAGGCTTACAGACAGTGAAGAAATGTTGCTGCGCGGCTTAAAGAAAATCTGTG 60
DB 15 AAAGTAAAGGCTTACAGACAGTGAAGAAATGTTGCTGCGCGGCTTAAAGAAAATCTGTG 74
QY 61 GTACCAACCCGAGGCGTTGAGAGAGCCCACTCCAGCTTCTTAAAGAGAGGTGCA 120
DB 75 GTACCAACCCGAGGCGTTGAGAGAGCCCACTCCAGCTTCTTAAAGAGAGGTGCA 134
QY 121 GGAAGTAAAGGCTTACAGACAGTGAAGAAATGTTGCTGCGCGGCTTAAAGAAAATCTGTG 180
DB 135 GGAAGTAAAGGCTTACAGACAGTGAAGAAATGTTGCTGCGCGGCTTAAAGAAAATCTGTG 194
QY 181 GCGGCTT 187
DB 195 GCGGCTT 201

RESULT 7
US-10-187-657-3
Sequence 3, Application US/10187657
Publication No. US20030068311A1
GENERAL INFORMATION:
APPLICANT: Laese, Amy K.W.
APPLICANT: Baughn, Mariah R.
APPLICANT: Azimzai, Yalda
TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
FILE REFERENCE: PV-0009 CIP
CURRENT APPLICATION NUMBER: US/10/187,657
PRIORITY FILING DATE: 2002-07-01
PRIORITY APPLICATION NUMBER: PCT/US00/07817
PRIORITY FILING DATE: 2000-03-22
PRIORITY APPLICATION NUMBER: 60/139,565
PRIORITY FILING DATE: 1999-06-16
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PERL Program
SEQ ID NO 3
LENGTH: 346
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20030068311A1 1752794F6
NAME/KEY: unsure
LOCATION: 91, 188, 206, 291, 337
OTHER INFORMATION: a, t, c, g, or other
US-10-187-657-3

Query Match 9.4%; Score 136; DB 14; Length 346;
Best Local Similarity 99.5%; Pred. No. 4.5e-56;
Matches 186; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 927 CTTGGCTTCATCTTACAGGCTGCGGCTTCCGCTTCAAGCTCTCTCCGCGCTAC 986
DB 1 CTTGGCTTCATCTTACAGGCTGCGGCTTCCGCTTCAAGCTCTCTCCGCGCTAC 60
QY 987 CACTAGTACGAGGCGGCGCTTCTGAGTACAGTGAAGAGCCGAGCTGCTGCTTCT 1046
DB 61 CACTAGTACGAGGCGGCGCTTCTGAGTACAGTGAAGAGCCGAGCTGCTGCTTCT 120
QY 1047 CGAGAGGCGGCTGAGTGTCTCAAGTATGTTGAGCCCAAGCGCTTCTGAGCCCTTCTGGA 1106
DB 121 CGAGAGGCGGCTGAGTGTCTCAAGTATGTTGAGCCCAAGCGCTTCTGAGCCCTTCTGGA 180
QY 1107 CCAAGC 1113
DB 181 CCAAGC 187

RESULT 8
US-10-187-657-7

```

FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(439)
OTHER INFORMATION: n = A,T,C or G
US-10-027-632-91970

Query Match
Best Local Similarity 100.0%; Score 107; DB 15; Length 439;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db
892 ACCACCGAGAGCTTCGCGCTCTTTCGGGGCTTCGCGCTTCACATCTCTAGACGCG 951
214 ACCACCGAGAGCTTCGCGCTCTTTCGGGGCTTCGCGCTTCACATCTCTAGACGCG 273

QY
952 CTCCTCCCGCTTCGCGCTTCGCGCTTCGCGCTTCACATCTCTAGACGCG 998
274 CTCCTCCCGCTTCGCGCTTCGCGCTTCGCGCTTCACATCTCTAGACGCG 320

RESULT 10
US-10-106-698-3277
; Sequence 3277; Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Riden et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypep
; FILE REFERENCE: PA005PI
; CURRENT APPLICATION NUMBER: US/10/106,698
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: Patentin Ver. 3.0
; SEQ ID NO 3277
; LENGTH: 512
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION (392)..(392)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION (453)..(453)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION (461)..(461)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION (493)..(493)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION (499)..(499)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION (509)..(509)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-3277

Query Match
Best Local Similarity 100.0%; Score 59; DB 14; Length 512;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY
405 GCGCTGATTTGGTGTGAGAGTTCTTCAGTCTGTTCATAGCGCGAGAAATTGTGG 463
220 GCGCTGATTTGGTGTGAGAGTTCTTCAGTCTGTTCATAGCGCGAGAAATTGTGG 278

RESULT 11
US-10-187-657-6

```

```
Sequence 6, Application US/10187657
Publication No. US20030068311A1
GENERAL INFORMATION:
APPLICANT: Laeek, Amy K.W.
APPLICANT: Baughn, Mariah R.
APPLICANT: Azimzai, Yalda
TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
FILE REFERENCE: PV-0009 CIP
CURRENT APPLICATION NUMBER: US/10/187,657
CURRENT FILING DATE: 2002-07-01
PRIOR APPLICATION NUMBER: PCT/US00/07817
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60/139,565
PRIOR FILING DATE: 1999-06-16
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PERL Program
SEQ ID NO 6
LENGTH: 506
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20030068311A1 SEQ#03652D1
FEATURE:
NAME/KEY: unsure
LOCATION: 87, 89, 93, 95, 124, 178, 227, 253, 262, 279, 285, 287, 293-294, 298,
LOCATION: 300, 303, 312, 319-321, 323, 325, 327, 332-333, 336, 343-344, 349, 377, 39
LOCATION: 405, 409-410, 412, 415, 420, 423, 429, 432, 436-437, 444-445, 448, 452, 45
LOCATION: 466, 469-471, 479, 483, 485, 492, 494, 499, 501, 504
OTHER INFORMATION: a, t, c, g, or other
US-10-187-657-6

Query Match          3.9%; Score 56; DB 14; Length 506;
Best Local Similarity 100.0%; Pred. No. 5.4e-17;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 479 CAGAAATGTTCTGGTGGTACGATGACACCAACATCCCTCAAGCCTCAGCGCA 534
DB 31 CAGAAATGTTCTGGTGGTACGATGACACCAACATCCCTCAAGCCTCAGCGCA 86

RESULT 12
US-10-424-599-64755
Sequence 64755, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 64755
LENGTH: 1398
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_29487C.1
US-10-424-599-64755

Query Match          2.6%; Score 38; DB 12; Length 1398;
Best Local Similarity 100.0%; Pred. No. 3.1e-08;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1404 TTTGTTTTTAAAAA
DB 1325 TTTGTTTTTAAAAA 1362
```

```
RESULT 13
US-10-116-255-18
Sequence 18, Application US/10116255
Publication No. US20030036646A1
GENERAL INFORMATION:
APPLICANT: Ni et al.
TITLE OF INVENTION: Uncoupling Protein Polynucleotides, Polypeptides, and
FILE REFERENCE: PT009P1
CURRENT APPLICATION NUMBER: US/10/116,255
CURRENT FILING DATE: 2002-04-05
PRIOR APPLICATION NUMBER: 09/685,897
PRIOR FILING DATE: 2000-10-11
PRIOR APPLICATION NUMBER: PCT/US00/09534
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: 60/128,701
PRIOR FILING DATE: 1999-04-09
PRIOR APPLICATION NUMBER: 60/142,821
PRIOR FILING DATE: 1999-07-08
PRIOR APPLICATION NUMBER: 60/149,448
PRIOR FILING DATE: 1999-08-18
PRIOR APPLICATION NUMBER: 60/164,751
PRIOR FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 66
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 18
LENGTH: 2165
TYPE: DNA
ORGANISM: Homo sapiens
US-10-116-255-18

Query Match          2.6%; Score 38; DB 14; Length 2165;
Best Local Similarity 100.0%; Pred. No. 3e-08;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1404 TTTGTTTTTAAAAA
DB 2128 TTTGTTTTTAAAAA 2165

RESULT 14
US-10-424-599-75074
Sequence 75074, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 75074
LENGTH: 3095
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_38805C.1
US-10-424-599-75074

Query Match          2.6%; Score 38; DB 12; Length 3095;
Best Local Similarity 100.0%; Pred. No. 2.9e-08;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1404 TTTGTTTTTAAAAA
DB 3047 TTTGTTTTTAAAAA 3084

RESULT 15
```

```
US-10-311-455-842
; Sequence 842, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; TITLE OF INVENTION: Cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 842
; LENGTH: 5642
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURES:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-842

Query Match      2.6%; Score 38; DB 14; Length 5642;
Best Local Similarity 100.0%; Pred. No. 2.8e-08;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY      1404 TTGCTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1441
          |||||||
DB      1374 TTGCTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1411
          |||||||
```

Search completed: February 24, 2004, 03:36:57
Job time : 557.173 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 23:20:25 ; Search time 3666.04 Seconds

(without alignments)
11737.841 Million cell updates/sec

Title: US-09-936-456-1
Sequence: 1441
aaagtaacgagctacagacag.....aaaaaaaaaaaaaaaa 1441

Scoring table: OLIGO-NTC
Gapop 60.0, Gapext 60.0

Searched: 27513289 segs, 14531090276 residues

Word size: 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

EST.*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estlun:*
5: em_estluc:*
6: em_estlpl:*
7: em_estlro:*
8: em_estlrc:*
9: gd_estl1:*
10: gd_estl2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_man:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phs:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	528	36.6	537	9	AI821606
2	515	35.7	510	9	AI826629
3	510	35.4	510	9	AI304327
4	446	31.0	446	9	AI984141

5	434	30.1	434	9	AI991272
6	414	28.7	414	9	AA593860
7	405	28.1	405	9	AI274929
8	389	27.0	389	9	AI791844
9	388	26.9	388	9	AI60493
10	382	26.5	382	9	AI983793
11	381	26.4	381	9	AI281211
12	375	26.0	375	14	CB854121
13	374	26.0	374	9	AI346155
14	369	25.6	369	9	AI60560
15	348	24.1	348	9	CB305399
16	330	22.9	330	9	AA573825
17	317	22.0	317	12	BM987789
18	296	20.5	296	9	AI821178
19	291	20.2	291	9	AM050605
20	280	19.4	280	12	BM977010
21	277	19.2	277	9	AI732165
22	270	18.7	270	14	CB850544
23	260	18.0	260	9	AM009962
24	252	17.5	252	9	AI262416
25	185	12.8	185	9	AI924216
26	183	12.7	183	12	BM973444
27	166	11.5	166	9	AA618335
28	149	10.3	149	9	AA469031
29	94	6.5	602	12	BO017315
30	88	6.1	151	9	AI695625
31	88	6.1	378	10	BF917041
32	44	3.1	119	9	AI921289
33	42	2.9	942	14	CF241985
34	41	2.8	279	9	AI270429
35	40	2.8	175	14	CF316906
36	40	2.8	217	14	CF328864
37	39	2.7	188	12	BM867036
38	39	2.7	225	9	AI932794
39	39	2.7	243	13	BO394290
40	39	2.7	248	9	AI537827
41	39	2.7	250	14	CD421839
42	39	2.7	284	9	AI537819
43	39	2.7	300	9	AI475331
44	39	2.7	324	10	AM130187
45	39	2.7	328	13	BU055639

ALIGNMENTS

RESULT 1
AI821606/c
LOCUS
DEFINITION
AI821606
VERSION
KEYWORDS
ORGANISM
SOURCE
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 537)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Other ESTs: nk08a1.y5
Contact: Robert Strauberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Stratagene, Inc., David B. Krizman, Ph.D.
CDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center

QY 1262 GCAGCCCGGAGACAGTCCCGGAGCCAGAGAGCTCCAGAAAGGACATGA 1321
 DB 155 GCAGCCCGGAGACAGTCCCGGAGCCAGAGAGCTCCAGAAAGGACATGA 96
 QY 1322 GCGCTGCTGGCGCGAGGCTCGGACATCCGAGGACACAGGAAAGTCTCTGCGCGCAT 1381
 DB 95 GCGCTGCTGGCGCGAGGCTCGGACATCCGAGGACACAGGAAAGTCTCTGCGCGCAT 36
 QY 1382 CTGTAATAAATTAACCTTTTCTTTGTTTTTAAA 1416.
 DB 35 CTGTAATAAATTAACCTTTTCTTTGTTTTTAAA 1

RESULT 3
 A1304327/c 510 bp mRNA linear EST 01-FEB-1999
 LOCUS 9057506.X1 NCI CGAP C08 Homo sapiens cDNA clone IMAGE:1912667 3'
 DEFINITION similar to contains TAR1.t3 TAR1 repetitive element // mRNA
 sequence.
 ACCESSION A1304327 GI:3988016
 VERSION A1304327.1
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 510)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-rt@mail.nih.gov
 Tissue Procurement: Christopher Moskalkuk, M.D., Ph.D., Michael R.
 Emert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/dbtp/image/image.html
 Insert length: 1487 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 476.
 Location/Qualifiers
 1..510
 source
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1912667"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B"
 /clone_1lb="NCI CGAP C08"
 /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 colon adenocarcinoma, and was then primed with a Not I -
 oligo(dT) primer. Double-stranded cDNA was ligated to Eco
 RI adaptors (Pharmacia), digested with Not I and cloned
 into the Not I and Eco RI sites of the modified pT73
 vector. Library is normalized. Library was constructed by
 Bento Soares and M. Fatima Bonaldo."

ORIGIN
 Query Match 35.4%; Score 510; DB 9; Length 510;
 Best Local Similarity 100.0%; Pred. No. 6.2e-92;
 Matches 510; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 904 TTGGCGCTTTGGGGGCTTCGGCTTGGCTTCATCTTCAAGGTGCGCGCTTGGCGGCTC 963
 DB 510 TTGGCGCTTTGGGGGCTTCGGCTTGGCTTCATCTTCAAGGTGCGCGCTTGGCGGCTC 451
 QY 964 CGCCTAAGCTCTCCGGCTCACCACTAGTACGCGCGCTTGGGTGACGCTGACA 1023

DB 450 CGCCTAAGCTCTCCGGCTCACCACTAGTACGCGCGCTTGGGTGACGCTGACA 391
 QY 1024 ACCGGGCTCTGTGCTCTTCTCTGGAGAGGCGGTGGTGAATCTCCAGTATGTCGCGCC 1083
 DB 390 ACCGGGCTCTGTGCTCTTCTCTGGAGAGGCGGTGGTGAATCTCCAGTATGTCGCGCC 331
 QY 1084 AGGCTCTTCGACCTTCTGGACCAAGCCCAAGACTGACCCAGAGAGAGAGGAGG 1143
 DB 330 AGGCTCTTCGACCTTCTGGACCAAGCCCAAGACTGACCCAGAGAGAGAGGAGG 271
 QY 1144 TCACCTCTTATCTCTGGAGACCCACTGACAGAGGCGCTCTCCAGACTTAAATGT 1203
 DB 270 TCACCTCTTATCTCTGGAGACCCACTGACAGAGGCGCTCTCCAGACTTAAATGT 211
 QY 1204 ATCACCACTTAACCTGTGAGGAGGAGCCCAATCTGACCTCTCCGCGCTTGGAGCATGCG 1263
 DB 210 ATCACCACTTAACCTGTGAGGAGGAGCCCAATCTGACCTCTCTCCGCGCTTGGAGCATGCG 151
 QY 1264 AGGCGGGAAGAGTGCCTGCGGCGGCTGCGCCAGAGAGTCCAGAGAGGCACTGAGC 1323
 DB 150 AGGCGGGAAGAGTGCCTGCGGCGGCTGCGCCAGAGAGTCCAGAGAGGCACTGAGC 91
 QY 1324 GCTGCTGGCGGAGGCTCGGACATCGCAGAGACCAAGAAAGTCTCTGCGGCGATCT 1383
 DB 90 GCTGCTGGCGGAGGCTCGGACATCGCAGAGACCAAGAAAGTCTCTGCGGCGATCT 31
 QY 1384 GTAAATTAACCTTTTCTTTGTTTTT 1413
 DB 30 GTAAATTAACCTTTTCTTTGTTTTT 1

RESULT 4
 A1984141/c 446 bp mRNA linear EST 27-OCT-1999
 LOCUS WU21C02.X1 Soares Dieckgraebe_colon_NHCD Homo sapiens cDNA clone
 DEFINITION IMAGE:2520674 3', mRNA sequence.
 ACCESSION A1984141
 VERSION A1984141.1 GI:5811360
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 446)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-rt@mail.nih.gov
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -40UP from Gibco.
 Location/Qualifiers
 1..446
 source
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2520674"
 /tissue_type="colonic mucosa from 3 patients with Crohn's
 disease"
 /lab_host="DH10B (phage-resistant)"
 /clone_1lb="Soares Dieckgraebe_colon_NHCD"
 /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 TGTACCAATCGAGCGGAGCGCGCGCGCTTTTCTTTTCTTTT 3']
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization. Tissue samples
 provided by Dr. Brian Dieckgraebe (Washington University,

```

//tissue_type="colonic mucosa from 3 patients with Crohn's
disease"
//lab_host="DH10B (phage-resistant)"
//clone_id="Scars Dieckgreife c773D_NHCD"
//note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
end of DNA derived with Not I; 2nd end derived with

```

```
/organism="Homo sapiens"  
/mol_type="mRNA"  
/db_xref="taxon:9606"  
/clone="IMAGE:2522647"
```

REFERENCE
AUTHORS
TITLE

1. (cases 1 to 4,4)

NCI-CCAP <http://www.ncbi.nlm.nih.gov/nciccap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL
COMMENT

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Stratagene, Inc.
DNA Sequencing by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1204 Std Error: 0.00
Seq primer: -40ml3 fwd. Et from Amersham
High quality sequence stop: 406.
Location/Qualifiers

FEATURES
source

1. .414
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1084349"
/sex="mixed"
/tissue_type="colon tumor"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: colon; Vector: Bluescript SK-; Site 1:
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
0190 dt. Pooled colon tumors. 5' adaptor sequence: 5'
GAATTCGGACGAG 3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.2 kb."

ORIGIN

Query Match 28.7%; Score 414; DB 9; Length 414;
Best Local Similarity 100.0%; Pred. No. 6.7e-73; Indels 0; Gaps 0;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 999 CGCGGCTTCTGGGTCAAGTGGCAACCGGCTCTGCTCTTCTCGAGAGGCGCT 1058
DB 414 CGCGGCTTCTGGGTCAAGTGGCAACCGGCTCTGCTCTTCTCGAGAGGCGCT 355
QY 1059 GGTAGTCTCCAGTATGTTGGCCGACGCTCTTTCGACCTTCTGACCAAGCCCAA 1118
DB 354 GGTAGTCTCCAGTATGTTGGCCGACGCTCTTTCGACCTTCTGACCAAGCCCAA 295
QY 1119 GACTGACGACGAGAGAGGGGCTCACTCTTATCTCTGCGAGCCCACTGACAGCA 1178
DB 294 GACTGACGACGAGAGAGGGGCTCACTCTTATCTCTGCGAGCCCACTGACAGCA 235
QY 1179 GAGCGCTTCTCCAGCTTAAATGTAACCACTAACCTGTGAGGGGAGCCCACTTGA 1238
DB 234 GAGCGCTTCTCCAGCTTAAATGTAACCACTAACCTGTGAGGGGAGCCCACTTGA 175
QY 1239 CTCTTCTCCGCGCTTGGGACATCGACGCGGAGAGAGTCCGCGAGGCTCGGCGCAG 1298
DB 174 CTCTTCTCCGCGCTTGGGACATCGACGCGGAGAGAGTCCGCGAGGCTCGGCGCAG 115
QY 1299 GAGAGCTCCAGAGAGGCACTGAGCGCTGTGGCGGAGGCTTGGACATCCGACGAC 1358
DB 114 GAGAGCTCCAGAGAGGCACTGAGCGCTGTGGCGGAGGCTTGGACATCCGACGAC 55
QY 1359 CAGGGAAGTCTCTCGGGGCGATCTGTAATAAATCTTTTCTTTTGTGTTTT 1412
DB 54 CAGGGAAGTCTCTCGGGGCGATCTGTAATAAATCTTTTCTTTTGTGTTTT 1

RESULT 7
A1274929/c 405 bp mRNA linear EST 29-JAN-1999
LOCUS A1274929
DEFINITION g149c11.x1 NCI CGAP Co8 Homo sapiens cDNA IMAGE:1875668 3'
similar to contains TARI.1c3 TARI repetitive element ;, mRNA
ACCESSION A1274929
VERSION A1274929.1 GI:3897203
KEYWORDS EST.

SOURCE

ORGANISM

Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 405)
REFERENCE
AUTHORS
TITLE
Natl Cancer Inst. Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1458 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 395.
Location/Qualifiers

FEATURES
source

1. .405
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1875668"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: pRT3D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from a
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pRT3
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 28.1%; Score 405; DB 9; Length 405;
Best Local Similarity 100.0%; Pred. No. 4.1e-71; Indels 0; Gaps 0;
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1009 TGGGTCAAGTGGCAACCGGCTCTGCTCTTCTCGAGAGGCGCTGTAAGTTC 1068
DB 405 TGGGTCAAGTGGCAACCGGCTCTGCTCTTCTCGAGAGGCGCTGTAAGTTC 346
QY 1069 CAGTATGTTGGCCGACGCTCTTTCGACCTTCTGACCAAGCCCAAGACTGACAC 1128
DB 345 CAGTATGTTGGCCGACGCTCTTTCGACCTTCTGACCAAGCCCAAGACTGACAC 286
QY 1129 CAGAGAGAGGGGCTCACTCTTATCTCTGAGACCCCACTGACCAAGAGGCGCTTC 1188
DB 285 CAGAGAGAGGGGCTCACTCTTATCTCTGAGACCCCACTGACCAAGAGGCGCTTC 226
QY 1189 CCAAGCTTAAATGTAACCACTTAACCTGTGAGGGGAGCCCAATCTGACTCTTCCC 1248
DB 225 CCAAGCTTAAATGTAACCACTTAACCTGTGAGGGGAGCCCAATCTGACTCTTCCC 166
QY 1249 GCCTTGGAGATCGCAGGCGGAGAGAGTCCCGCAGGCTGGGCGGAGAGTCCA 1308
DB 165 GCCTTGGAGATCGCAGGCGGAGAGAGTCCCGCAGGCTGGGCGGAGAGTCCA 106
QY 1309 GGAAGGCACTGAGCGCTGTGGCGGAGGCTTCGACATCCGACGACGAGAAAGT 1368
DB 105 GGAAGGCACTGAGCGCTGTGGCGGAGGCTTCGACATCCGACGAGGAAAGT 46
QY 1369 CTCTTGGGCGATCTGTAATAAATCTTTTCTTTTGTGTTTT 1413
DB 45 CTCTTGGGCGATCTGTAATAAATCTTTTCTTTTGTGTTTT 1

RESULT 8
 A1791844 505 bp mRNA linear EST 13-DEC-1999
 LOCUS nk08a11.y5 NCI_CGAP Co2 Homo sapiens cDNA clone IMAGE:1012892 5'
 DEFINITION similar to contains Alu repetitive element/, mRNA sequence.
 ACCESSION A1791844
 VERSION A1791844.1 GI:5339486
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 505)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Other ESTs: nk08a11.s1
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Stratagene, Inc., David B. Kitzman,
 Ph.D.
 CDNA Library Arraying: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html

This read is a RESEQUENCE of a previously sequenced human clone
 Original clone citation: see original entry for original citation
 Information
 This 5' resequenced clone has no previous 5' data to verify this
 new read against
 Putative full length read
 The vector to vector length is 671
 Insert Length: 741 Std Error: 0.00
 Seq primer: -40RP from Gibco
 High quality sequence stop: 418.
 Location/Qualifiers
 1..505
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1012892"
 /tissue_type="tumor"
 /lab_host="SOLR (kanamycin resistant)"
 /clone_lib="NCI CGAP Co2"
 /note="Organ: colon; Vector: Bluescript SK-; Site: 1;
 EcoRI; Site: 2: XhoI; Cloned unidirectionally. Primer:
 Oligo dr. Bulk colon villous adenoma. 5' adaptor sequence:
 5' GAATTCGGACGAG 3' 3' adaptor sequence: 5'
 CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 Kb."

ORIGIN
 Query Match 27.0%; Score 389; DB 9; Length 505;
 Best Local Similarity 100.0%; Pred. No. 4.8e-68;
 Matches 389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 736 AAGTTACACCGAGTAGCCCTTGGGGCTGTACACACAGTACCACTGGCGGAGACATAC 795
 DB 1 AAGTTACACCGAGTAGCCCTTGGGGCTGTACACACAGTACCACTGGCGGAGACATAC 60
 QY 796 GCTTCGGCGACGCTATAGGTGGGCTTCTGCTTGGCTCTCTCCAAAGTGCCTCTCC 855
 DB 61 GCTTCGGCGACGCTATAGGTGGGCTTCTGCTTGGCTCTCTCCAAAGTGCCTCTCC 120
 QY 856 ACGCGGCGCGCTCTACGAGAGGCTGTGCACTGTGACCAACCGGAGCTTTCGCGCTTTC 915
 DB 121 ACGCGGCGCGCTCTACGAGAGGCTGTGCACTGTGACCAACCGGAGCTTTCGCGCTTTC 180

RESULT 9
 A1660493/c 538 bp mRNA linear EST 18-DEC-1999
 LOCUS w67h02.x1 Soares_Dieckgraefe_colon_NHCD Homo sapiens cDNA clone
 DEFINITION IMAGE:2346195 3' similar to contains TARK1.t3 TARK1 repetitive
 element/, mRNA sequence.
 ACCESSION A1660493
 VERSION A1660493.1 GI:4764063
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 538)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 676 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 463.
 Location/Qualifiers
 1..538
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2346195"
 /tissue_type="colonic mucosa from 3 patients with Crohn's
 disease"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="Soares_Dieckgraefe_colon_NHCD"
 /note="Organ: colon; Vector: pTV73D-Pac (Pharmacia) with a
 modified polylinker; Site: 1: Not I; Site 2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 TGTACCAATCTGAGTGGAGCGCGCCGCTCTTTTCTTTTTTTTTTTT 3']
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pTV73 vector. Library
 went through one round of normalization. Tissue samples
 provided by Dr. Brian Dieckgraefe (Washington University,
 dieckim.wustl.edu); colonic mucosa represents a range of
 disease involvement from moderate to severe Crohn's
 disease; samples include both perforating (fistulas) and
 non-perforating samples. Library constructed by Bento
 Soares and M. Fatima Bonaldo."

ORIGIN
 Query Match 26.9%; Score 388; DB 9; Length 538;
 Best Local Similarity 100.0%; Pred. No. 7.1e-68;
 Matches 388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1027 GAGCTCTTGCTCTCTCGAGAGGCGCGTGAAGTCTCCAGTATTCGAGCCACAGC 1086

```

Db      388 GGGCTCTGTGCTCTCTCTCTCGAGAGGGCCGTGTGATCTCCAGTATGTTTGGCCCAAC 329
QY      1087 GCTCTTGGACACCTTCTTGACCAAGGCGCCAGAGCTGAGCCAGAGAGAGGGGCTCA 1146
Db      328 GCTCTTGGACACCTTCTTGACCAAGGCGCCAGAGCTGAGCCAGAGAGAGGGGCTCA 269
QY      1147 CCTCTTATCCCTCGGCGACCACTGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1206
Db      268 CCTTATCTCTGCGACCACTGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATC 209
QY      1207 ACCACTAACCTGTGAGGGAGGACCAATCTGACTCTTCCCGCTTGGAGACATCGAAG 1266
Db      208 ACCACTAACCTGTGAGGGAGGACCAATCTGACTCTTCCCGCTTGGAGACATCGAAG 149
QY      1267 CCGGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1326
Db      148 CCGGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 89
QY      1327 GCTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1386
Db      88 GCTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 29
QY      1387 AATAACCTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 1414
Db      28 AATAACCTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 1

```

```

RESULT 10
LOCUS   A1983793/c 411 bp mRNA linear EST 27-OCT-1999
DEFINITION  wu20c09.x1 Soares, Dieckgraefe_colon_NHCD Homo sapiens CDNA clone
VERSION   A1983793
XREFS     IMAGE:2520592 3, mRNA sequence.
KEYWORDS  A1983793.1 GI:5811012
SOURCE    EST.
ORGANISM  Homo sapiens (human)
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 411)
          NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
          National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
          Tumor Gene Index
          Unpublished (1997)
          Contact: Robert Strausberg, Ph.D.
          Email: cgaps-remail.nih.gov
          This clone is available royalty-free through LIND; contact the
          IMAGE Consortium (info@image.llnl.gov) for further information.
          Seq primer: -40UP from Gdbco
          High quality sequence stop: 400.
          Location/Qualifiers
            1..411
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="IMAGE:2520592"
              /tissue_type="colonic mucosa from 3 patients with Crohn's
              disease"
              /lab_host="DH10B (phage-resistant)"
              /clone_id="Soares_Dieckgraefe_colon_NHCD"
              /note="Organ: Colon; Vector: pT73D-Pac (Pharmacia) with a
              modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
              strand CDNA was primed with a Not I - oligo(dt) primer [5'
              TGTACCAATCTGAAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
              double-stranded CDNA was ligated to Eco RI adaptors
              (Pharmacia), digested with Not I and cloned into the Not I
              and Eco RI sites of the modified pT73 vector. Library
              went through one round of normalization. Tissue samples
              provided by Dr. Brian Dieckgraefe (Washington University,
              dieck@im.wustl.edu); colonic mucosa represents a range of
              disease involvement from moderate to severe Crohn's
              disease; samples include both perforating (fistulas) and

```

ORIGIN

non-perforating samples. Library constructed by Bento Soares and M. Fatima Bonaldo.

Query Match 26.5%; Score 382; DB 9; Length 411;
 Best Local Similarity 100.0%; Pred. No. 1.4e-66;
 Matches 382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      1008 CTGGGTCAGGCTGGCAACCGGCGTCTGCTGCTCTCTCTCGAGAGGGCCGTGTGAGTCT 1067
Db      411 CTGGGTCAGGCTGGCAACCGGCGTCTGCTGCTCTCTCTCTCGAGAGGGCCGTGTGAGTCT 352
QY      1068 CCAATATGTTGGCCCGGAGGCTCTTTCGACACCTTTCGACCAAGGCGCAAGAGACTGAG 1127
Db      351 CCAATATGTTGGCCCGGAGGCTCTTTCGACACCTTTCGACCAAGGCGCAAGAGACTGAG 292
QY      1128 CCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1187
Db      291 CCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 232
QY      1188 CCGAGACTTAAATGTATGATACCACTAATCTGTGAGGGAGACCAATCTGACTCTTCC 1247
Db      231 CCGAGACTTAAATGTATGATACCACTAATCTGTGAGGGAGACCAATCTGACTCTTCC 172
QY      1248 CGCCTTGGACATCGAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1307
Db      171 CGCCTTGGACATCGAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 112
QY      1308 AGGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1367
Db      111 AGGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 52
QY      1368 TCTCTGGGCGGATCTGTAAT 1389
Db      51 TCTCTGGGCGGATCTGTAAT 30

```

```

RESULT 11
LOCUS   A1281211/c 381 bp mRNA linear EST 28-JAN-1999
DEFINITION  qk58e08.x1 NCI CGAP C08 Homo sapiens CDNA clone IMAGE:1873190 3
VERSION   A1281211
XREFS     IMAGE:1873190 3, mRNA sequence.
KEYWORDS  A1281211.1 GI:3919444
SOURCE    EST.
ORGANISM  Homo sapiens (human)
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 381)
          NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
          National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
          Tumor Gene Index
          Unpublished (1997)
          Contact: Robert Strausberg, Ph.D.
          Email: cgaps-remail.nih.gov
          Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
          Emmerit-Buck, M.D., Ph.D.
          CDNA Library Preparation: M. Bento Soares, Ph.D.
          CDNA Library Arrayed by: Greg Lennon, Ph.D.
          DNA Sequencing by: Washington University Genome Sequencing Center
          clone distribution: NCI-CGAP clone distribution information can be
          found through the I.M.A.G.E. Consortium/LIND at:
          www-bio.llnl.gov/dbp/image/image.html
          Insert Length: 1422 Std Error: 0.00
          Seq primer: -40UP from Gdbco
          High quality sequence stop: 340.
          Location/Qualifiers
            1..381
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"

```

FEATURES

source

ORIGIN

/clone="IMAGE:1873190"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B"
 /clone_id="NCI CGAP C08"
 /note="Organ: colon; Vector: pT73-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from colon adenocarcinoma, and was then primed with a Not I - oligo (dT) primer. Double-stranded cDNA was ligated to Eco RI adaptor (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

Query Match 26.4%; Score 381; DB 9; Length 381;
 Best Local Similarity 100.0%; Pred. No. 2,4e-66;
 Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1033 CTGTGCTCTCTCTCTGAGAGGCGCTGTGATCTCCAGTATGTTGCGCCCGACGCTCTT 1092
 |||||
 381 CTGTGCTCTCTCTCTGAGAGGCGCTGTGATCTCCAGTATGTTGCGCCCGACGCTCTT 322
 |||||
 1093 CGCACCCCTTCTGACCAAGCGCCCAAGACTCGAGGCGCAGAGAGGGGGCTCACTCTT 1152
 |||||
 321 CGACCCCTTCTGACCAAGCGCCCAAGACTCGAGGCGCAGAGAGGGGGCTCACTCTT 262
 |||||
 1153 ATCTCGGCGACCACTGACCAAGAGGCGCTCTCCAGACTTAAATGTATCACCCT 1212
 |||||
 261 ATCTCGGCGACCACTGACCAAGAGGCGCTCTCCAGACTTAAATGTATCACCCT 202
 |||||
 1213 AACCTGTAGGGGGAGCCCAATCTGATCTCTTCCCGCTTGGGACATCGAGGCGCGGA 1272
 |||||
 201 AACCTGTAGGGGGAGCCCAATCTGATCTCTTCCCGCTTGGGACATCGAGGCGCGGA 142
 |||||
 1273 AGCAGTCCCGCCGACGCTGAGCGCCAGAGAGCTCCAGAGAGGCGCTGAGCGCTGAGC 1332
 |||||
 141 AGCAGTCCCGCCGACGCTGAGCGCCAGAGAGCTCCAGAGAGGCGCTGAGCGCTGAGC 82
 |||||
 1333 GCGAGGCTCTGACATCGGACGAGGACCAAGGAAATCTCTGAGGCGCATCTGTAAATAAA 1392
 |||||
 81 GCGAGGCTCTGACATCGGACGAGGACCAAGGAAATCTCTGAGGCGCATCTGTAAATAAA 22
 |||||
 1393 CCTTTTCTTTTCTTTT 1413
 |||||
 21 CCTTTTCTTTTCTTTT 1

RESULT 12

CB854121 573 bp mRNA linear EST 22-APR-2003
 LOCUS CB854121/c
 DEFINITION UT-CF-DUI-aal-1-16-0-UI s1 UT-CF-DUI Homo sapiens cDNA clone

ACCESSION CB854121
 VERSION CB854121.1 GI:30044498
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 573)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 PUBMED 8889548

COMMENT Contact: McCray, PB
 McCray Lab
 University of Iowa
 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
 Tel: 319 356 4866
 Fax: 319 356 7171
 Email: paul-mccray@uiowa.edu

FEATURES

source

Tissue Procurement: Dr. M. J. Welsh, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
 The following repetitive elements were found in this cDNA sequence: 533-573, >Alu (matched complement)
 Seq primer: M13 FORWARD
 POLYA=No.

Location/Qualifiers
 1..573
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-CF-DUI-aal-1-16-0-UI"
 /tissue_type="Primary Lung Epithelial Cells"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_id="UI-CF-DUI"
 /note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-CF-DUI is a normalized cDNA library containing the following tissue(s): Primary Lung Epithelial Cells The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GCGTGTAGCG.
 TAG_SBO=None found"

ORIGIN

Query Match 26.0%; Score 375; DB 14; Length 573;
 Best Local Similarity 100.0%; Pred. No. 2,4e-65;
 Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1044 CCTCGAGAGGGCGCTGTGATCTCCAGTATGTTGCGCCCGACGCTCTTGGACCCCTCT 1103
 |||||
 482 CCTCGAGAGGGCGCTGTGATCTCCAGTATGTTGCGCCCGACGCTCTTGGACCCCTCT 423
 |||||
 1104 GGACCAAGCGCCCAAGACTGACGAGGAGAGGGGGCTCACTTATCTCGCGGA 1163
 |||||
 422 GGACCAAGCGCCCAAGACTGACGAGGAGAGGGGGCTCACTTATCTCGCGGA 363
 |||||
 1164 CCCTGTCACCAAGAGGCGCTCTTCCAGACTTAAATGTATCACCCTAAGCTGTAGG 1223
 |||||
 362 CCCTGTCACCAAGAGGCGCTCTTCCAGACTTAAATGTATCACCCTAAGCTGTAGG 303
 |||||
 1224 GGGACCCATCTGACCTCTTCCCGCTTGGACATCGAGGCGGGGAAAGCATGTGCCG 1283
 |||||
 302 GGGACCCATCTGACCTCTTCCCGCTTGGACATCGAGGCGGGGAAAGCATGTGCCG 243
 |||||
 1284 CCAGGCTTGGGCGCAGAGAGCTCCAGAGAGGCACTGAGCGCTGTGGCGGAGCCTTG 1343
 |||||
 242 CCAGGCTTGGGCGCAGAGAGCTCCAGAGAGGCACTGAGCGCTGTGGCGGAGCCTTG 183
 |||||
 1344 GACATCCGAGGACACAGGAAAGCTCTCTGGGGGAGATCTAATAAACCCTTTTCT 1403
 |||||
 182 GACATCCGAGGACACAGGAAAGCTCTCTGGGGGAGATCTAATAAACCCTTTTCT 123
 |||||
 1404 TTTGTTTTTAAAAA 1418
 |||||
 122 TTTGTTTTTAAAAA 108

RESULT 13

AI346155/c
 LOCUS AI346155 405 bp mRNA linear EST 02-FEB-1999

DEFINITION	9433f12.x1 NCI CGAP CoB Homo sapiens cDNA clone IMAGE:1925807 3'
ACCESSION	U03332
VERSION	1.1
KEYWORDS	EST
SOURCE	Human
ORGANISM	Homo sapiens (human)
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/MLN at: www.bio.lnlnl.gov/bbrr/image/image.html Insert Length: 1488 Std Error: 0.00 Seq primer: -400P from Gldpo High quality sequence stop: 375. Location/Qualifiers 1..405 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:1925807" /tissue_type="adenocarcinoma" /lab_host="DH10B" /clone_id="NCI CGAP CoB" /note="Torgan: colon; Vector: pMT23-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from colon adenocarcinoma, and was then primed with a Not I - 0150 (dr) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pMT23 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."
FEATURES	source
Query Match	26.0%; Score 374; DB 9; Length 405;
Best Local Similarity	100.0%; Pred. No. 5.4e-65;
Matches 374; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1009 TGGGTCACGCTGGCAACCGGCGCTCTGCGCTCTTCTCGAGAGGGGCGGTGTAGTCTC 10688
DB	405 TGGGTCACGCTGGCAACCGGCGCTCTGCGCTCTTCTCGAGAGGGGCGGTGTAGTCTC 346
QY	1069 CAGTATGTTCCGCCCAAGCGCTTTTGCACCCCTTCTGACCAAGGCGCCAGAGCTGCAGC 11288
DB	345 CAGTATGTTCCGCCCAAGCGCTTTTGCACCCCTTCTGACCAAGGCGCCAGAGCTGCAGC 286
QY	1129 CAGGAGAGAGGGGGCTCAGCTCTTATCTCCGGCAGACCCATGTCACAAAGAGCGCTCTC 11888
DB	285 CAGGAGAGAGGGGGCTCAGCTCTTATCTCCGGCAGACCCATGTCACAAAGAGCGCTCTC 226
QY	1189 CCAGACTTAATGTAATGATACCACTAACCTGTGAGGGGAGCCCAATCTGACTCTCTCC 1248
DB	225 CCAGACTTAATGTAATGATACCACTAACCTGTGAGGGGAGCCCAATCTGACTCTCTCC 166
QY	1249 GCGTTGGGACATTCGACAGGCGGGGAGACAGTGTGCGGCGAGGCTGGGCGCAGAGAGCTCA 13088
DB	165 GCGTTGGGACATTCGACAGGCGGGGAGACAGTGTGCGGCGAGGCTGGGCGCAGAGAGCTCA 106
QY	1309 GGAAGGGGACATGAGGCGCTCTGAGCGGAGAGGCTCGGACATTCGCGAGGACACAGGAGAGT 13688

[illegible]

QY 1189 CCAGACTTAAATGATGACCACTAACCTGTGAGGGGAGCCCAATCTGACTCTTCC 1248
 Db 240 CCAGACTTAAATGATGACCACTAACCTGTGAGGGGAGCCCAATCTGACTCTTCC 181
 QY 1249 GCGTTGGGACATGAGGCGGGGAGAGCATGCGCCGAGGCTGGGCGAGAGAGCTCCA 1308
 Db 180 GCGTTGGGACATGAGGCGGGGAGAGCATGCGCCGAGGCTGGGCGAGAGAGCTCCA 121
 QY 1309 GGAAGGGCACTGAGCGCTGCTGCGCGAGGCGCTGGACATCCGAGGCAACGAGGAAGT 1368
 Db 120 GGAAGGGCACTGAGCGCTGCTGCGCGAGGCGCTGGACATCCGAGGCAACGAGGAAGT 61
 QY 1369 CTCCTGGGGCGATCTGTAATTAATTAACCTTTTCTTTCTTTTAAATTAATTAATTA 1428
 Db 60 CTCCTGGGGCGATCTGTAATTAATTAACCTTTTCTTTCTTTTAAATTAATTAATTA 1

RESULT 15
 CB305399 737 bp mRNA linear EST 04-MAR-2003
 LOCUS
 DEFINITION UT-CF-EN1-aed-m-17-0-UI s1 UT-CF-EN1 Homo sapiens cDNA clone
 ACCESSION CB305399
 VERSION CB305399.1 GI:28845910
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 737)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 PubMed 8889548
 COMMENT Contact: McCray, PB
 McCray Lab
 University of Iowa
 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
 Tel: 319 356 4866
 Fax: 319 356 7171
 Email: paul.mccray@uiowa.edu
 Tissue Procurement: Dr. M. J. Welsh, University of Iowa
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com) or from Open Biosystems
 (www.openbiosystems.com).
 The following repetitive elements were found in this cDNA
 sequence: 29-138, >MIR#SINE/MIR (matched complement) 195-290, >ALU
 223-320, >ALU
 Seq primer: M13 FORWARD
 POLYA=Yes.

FEATURES

Source
 Location/Qualifiers
 1..737
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-CF-EN1-aed-m-17-0-UI"
 /tissue_type="Primary Lung Cystic Fibrosis Epithelial
 Cells"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies) (T1 phase resistant)"
 /clone_id="UI-CF-EN1"
 /note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I;
 UI-CF-EN1 is a normalized cDNA library containing the
 following tissue(s): Primary Lung Cystic Fibrosis
 Epithelial Cells. The library was constructed according to
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,

1996. First strand cDNA synthesis was primed with an
 oligo-dT primer containing a Not I site. Double stranded
 cDNA was ligated to an EcoR I adaptor, digested with Not
 I, and cloned directionally into pT73-Pac vector. The
 oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tag for this library is CTGCTCAGGT.
 TAG_TISSUE=Human Lung Epithelial Cell lines untreated LPS
 6hr to LPS 24h
 TAG_LIB=UI-CF-EN1
 TAG_SEQ=CTGCTCAGGT"

ORIGIN
 Query Match 24.1%; Score 348; DB 14; Length 737;
 Best Local Similarity 100.0%; Pred. No. 4,1e-60;
 Matches 348; Conservative 0; Mismatch 0; Indels 0; Gaps 0;
 QY 1027 GAGCGTCTGTGCTCTTCTCTCGAGAGGGCGGTGAGTCTCAGATATGTTGGCCGAC 1086
 Db 353 GAGCGTCTGTGCTCTTCTCTCGAGAGGGCGGTGAGTCTCAGATATGTTGGCCGAC 412
 QY 1087 GCTCTTGGCACTCTTCTGACCAAGAGGAGGAGTCTGAGGAGGAGGAGGAGGAGG 1146
 Db 413 GCTCTTGGCACTCTTCTGACCAAGAGGAGGAGTCTGAGGAGGAGGAGGAGGAGG 472
 QY 1147 CCTCTTATCTCTGCGGAGCCCACTGACAGAGGCGCTCTCCAGACTTAAATGATATC 1206
 Db 473 CCTCTTATCTCTGCGGAGCCCACTGACAGAGGCGCTCTCCAGACTTAAATGATATC 532
 QY 1207 ACCACTTAACCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCTTGGACATCCGAGG 1266
 Db 533 ACCACTTAACCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCTTGGACATCCGAGG 552
 QY 1267 CCGGAGAGCATGTCGCCGAGGCTGGGCGCGAGAGAGCTCCAGAGAGGAGGAGGAGCT 1326
 Db 593 CCGGAGAGCATGTCGCCGAGGCTGGGCGCGAGAGAGCTCCAGAGAGGAGGAGGAGCT 652
 QY 1327 GCTGGCGGAGGCTCTGACATCCGACAGGCAACGAGGAAGTCTCTCG 1374
 Db 653 GCTGGCGGAGGCTCTGACATCCGACAGGCAACGAGGAAGTCTCTCG 700

Search completed: February 24, 2004, 03:21:09
 Job time : 3670.04 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 19:34:53 ; Search time 5685.97 Seconds

(without alignment cell) updates/sec

Title: US-09-936-456-1

Perfect score: 1441

Sequence: 1 aaagtaacgctcagacag.....aaaaaaaaaaaaaaaa 1441

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenDbml:*
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: gb_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hlg_hum:*
31: em_hlg_inv:*
32: em_hlg_other:*
33: em_hlg_mus:*
34: em_hlg_dln:*
35: em_hlg_rod:*
36: em_hlg_mam:*
37: em_hlg_vrt:*
38: em_sy:*
39: em_hlgo_hum:*
40: em_hlgo_mus:*
41: em_hlgo_other:*

score greater than or equal to the score being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1441	100.0	1441	6 AX035346	AX035346 Sequence
2	1029	71.4	1740	9 HSM806241	BM337861 Homo sapi
3	711.4	49.4	1354	10 BC031111	BC031111 Mus muscu
4	626.2	43.5	5851	4 AF547266	AF547266 Sus scrofa
5	465.4	32.3	498	6 AX035348	AX035348 Sequence
6	405.4	28.1	16637	2 AC012255	AC012255 Homo sapi
7	405.4	28.1	181312	9 AC091117	AC091117 Homo sapi
8	403.8	28.0	156534	2 AC009700	AC009700 Homo sapi
9	403.8	28.0	171444	2 AC087790	AC087790 Homo sapi
10	396	27.5	2684	6 AX114511	AX114511 Sequence
11	396	27.5	2684	6 AK056896	AK056896 Homo sapi
12	396	27.5	156534	2 AC009700	AC009700 Homo sapi
13	356.4	24.7	1636	6 AX405919	AX405919 Sequence
14	356.4	24.7	1923	9 BC029819	BC029819 Homo sapi
15	348.8	24.2	1521	10 BC019755	BC019755 Mus muscu
16	289.4	24.3	2439	9 AK027187	AK027187 Homo sapi
17	245.8	17.1	1706	9 BC020841	BC020841 Homo sapi
18	245.8	17.1	2684	6 AX114511	AX114511 Sequence
19	245.8	17.1	2684	9 AK056896	AK056896 Homo sapi
20	217	15.1	217	9 H518312F	Z57952 H. sapiens C
21	199.8	13.9	218	9 H518312R	Z59954 H. sapiens C
22	167.8	11.6	229583	10 AL844566	AL844566 Mouse DNA
23	166	11.5	1923	9 BC029819	BC029819 Homo sapi
24	165.8	11.5	231868	2 AC118124	AC118124 Rattus no
25	159.4	11.1	166937	2 AC012255	AC012255 Homo sapi
26	159.4	11.1	171444	2 AC087790	AC087790 Homo sapi
27	159.4	11.1	181312	9 AC091117	AC091117 Homo sapi
28	154.2	10.7	231868	2 AC118124	AC118124 Rattus no
29	147	10.2	229583	10 AL844566	AL844566 Mouse DNA
30	123.2	8.5	110250	2 AC138220	AC138220 Mus muscu
31	121.6	8.4	55061	2 AC091597	AC091597 Mus muscu
32	101	7.0	110250	2 AC138220	AC138220 Mus muscu
33	65.6	4.6	2207	3 AY061346	AY061346 Drosophila
34	63.4	4.4	125020	9 AF429315	AF429315 Homo sapi
35	61	4.2	7218	6 I66494	I66494 Sequence 14
36	60.6	4.2	2000	6 AX655393	AX655393 Sequence
37	59.6	4.1	125020	9 AF429315	AF429315 Homo sapi
38	56.8	3.9	63689	2 AC024268	AC024268 Homo sapi
39	55.8	3.9	4615	1 STWHB12X	L22864 Streptomyces
40	55.2	3.8	151173	8 AC135257	AC135257 Genomic s
41	53	3.7	2690	8 AY332473	AY332473 Oryza sat
42	53	3.7	169162	8 AP004267	AP004267 Oryza sat
43	52	3.6	2000	6 AX655393	AX655393 Sequence
44	52	3.6	117787	2 AC104271	AC104271 Oryza sat
45	51.8	3.6	168866	8 OSUN00023	AL606368 Oryza sat

ALIGNMENTS

RESULT 1
LOCUS AX035346 1441 bp DNA linear PAT 15-NOV-2000
DEFINITION Sequence 1 from Patent WO0053748.
ACCESSION AX035346
VERSION AX035346.1 GI:1191064
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Bruck,C.E., Vinals,Y.D., Coche,T. and Cassart,J.P.
TITLE Novel compounds
JOURNAL Patent: WO 0053748-A 1 14-SEP-2000;

BRUCK CLAUDINE ELVIRE MARIE (BE) ; SMITHKLINE BEECHAM BIOLOG (BE) ;
VINLUS Y DE BASOLS CARLOTTA (BE) ; COCHE THIERRY (BE) ; CASPART
JEAN POL (BE)

FEATURES

Location/Qualifiers
1..1441
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 1441; DB 6; Length 1441;
Best Local Similarity 100.0%; Pred. No. 3,3e-294;
Matches 1441; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGTAAAGGCTTACAGACAGTAAATAGTTGGCTGGCGGCTAGAAAACTCTGCG 60
DB 1 AAGTAAAGGCTTACAGACAGTAAATAGTTGGCTGGCGGCTAGAAAACTCTGCG 60

QY 61 GTACCAACCCAGAGGCTTGAAGAGAGCCACCTCCAGCTTCTTAAAGAGAGTGA 120
DB 61 GTACCAACCCAGAGGCTTGAAGAGAGCCACCTCCAGCTTCTTAAAGAGAGTGA 120

QY 121 GAGTCAAGCTTCAACAGCCCACTGGTCCAGCTTGTAGCAAAAGAGAGCCAGAG 180
DB 121 GAGTCAAGCTTCAACAGCCCACTGGTCCAGCTTGTAGCAAAAGAGAGCCAGAG 180

QY 181 GCGCTCCGCGCTCCAGGAGCCAGCTTGGCTGGCTGGCGGCTGGAGAGC 240
DB 181 GCGCTCCGCGCTCCAGGAGCCAGCTTGGCTGGCTGGCGGCTGGAGAGC 240

QY 241 ACTCGGCGGCTGGAGAGTGAACCTGTGAAAGGCGATAGCTTTTAAAGAGAGC 300
DB 241 ACTCGGCGGCTGGAGAGTGAACCTGTGAAAGGCGATAGCTTTTAAAGAGAGC 300

QY 301 CGGATAGCGGCTTCAAGCTTCACTGTCTCATCGTATCTAGTGTGTTTGGCTCTA 360
DB 301 CGGATAGCGGCTTCAAGCTTCACTGTCTCATCGTATCTAGTGTGTTTGGCTCTA 360

QY 361 GAGCAAGGCTTCTGCTCATCTTGGCGGGATCCGTGGCATCTCGCGCTGGTTGG 420
DB 361 GAGCAAGGCTTCTGCTCATCTTGGCGGGATCCGTGGCATCTCGCGCTGGTTGG 420

QY 421 GTAGAGTCTTCTCAAGTCTGTCTATAGAGGAGAAATGGTGGCTGTGCACTTCA 480
DB 421 GTAGAGTCTTCTCAAGTCTGTCTATAGAGGAGAAATGGTGGCTGTGCACTTCA 480

QY 481 GAATGTTGTTGAGTACAGTGAACCAACCAACATCTTACAAAGCTTCAAGAGAG 540
DB 481 GAATGTTGTTGAGTACAGTGAACCAACCAACATCTTACAAAGCTTCAAGAGAG 540

QY 541 GTTACAGCCGCTGGCTGTGCTGTGGGCTGTGAGGAGGATTAATATTAATTA 600
DB 541 GTTACAGCCGCTGGCTGTGCTGTGGGCTGTGAGGAGGATTAATTAATTAATTA 600

QY 601 ACCCGAGTGCATGATGAGAGAGCAATGACTTACAGAGAGTTCACCTGGGCTG 660
DB 601 ACCCGAGTGCATGATGAGAGAGCAATGACTTACAGAGAGTTCACCTGGGCTG 660

QY 661 AAGAGAAATTAAGCGCGGAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
DB 661 AAGAGAAATTAAGCGCGGAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720

QY 721 CTCTACCTGGCGAGAGTTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
DB 721 CTCTACCTGGCGAGAGTTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780

QY 781 CTGGCGGAG 840
DB 781 CTGGCGGAG 840

QY 841 AAGTGTGCTCTTCAAGCGGCGGCTTCAAGAGAGAGAGAGAGAGAGAGAGAG 900
DB 841 AAGTGTGCTCTTCAAGCGGCGGCTTCAAGAGAGAGAGAGAGAGAGAGAGAG 900

QY 901 GCGTTCGGGCTCTTCCGAGGCTTTCGAGCTTCCATCTAGAGTGGCGCTGCGCG 960
DB 901 GCGTTCGGGCTCTTCCGAGGCTTTCGAGCTTCCATCTAGAGTGGCGCTGCGCG 960

QY 961 CTCCGCTTCCGCTCTCCGCTTCCAGTCAAGAGAGAGAGAGAGAGAGAGAGAG 1020
DB 961 CTCCGCTTCCGCTCTCCGCTTCCAGTCAAGAGAGAGAGAGAGAGAGAGAGAG 1020

QY 1021 GGAACCGGCGCTCTGCTCTCTTCTTCTGAGAGAGAGAGAGAGAGAGAGAGAG 1080
DB 1021 GGAACCGGCGCTCTGCTCTCTTCTTCTGAGAGAGAGAGAGAGAGAGAGAGAG 1080

QY 1081 CCAGAGCTCTTCCAGAGCTTCTGAGCAAAAGAGAGAGAGAGAGAGAGAGAGAG 1140
DB 1081 CCAGAGCTCTTCCAGAGCTTCTGAGCAAAAGAGAGAGAGAGAGAGAGAGAGAG 1140

QY 1141 GAGTCACTCTTATCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
DB 1141 GAGTCACTCTTATCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200

QY 1201 TGTATCACTTATCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
DB 1201 TGTATCACTTATCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260

QY 1261 CGCAGAGCGGAG 1320
DB 1261 CGCAGAGCGGAG 1320

QY 1321 AGCGCTGCTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
DB 1321 AGCGCTGCTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380

QY 1381 TCTGTAAATAACTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1440
DB 1381 TCTGTAAATAACTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1440

QY 1441 A 1441
DB 1441 A 1441

RESULT 2
HSM806241 1740 bp mRNA linear PRI 17-JUN-2003
LOCUS Homo sapiens mRNA: CDNA DKFZp686C04213 (from clone DKFZp686C04213).
DEFINITION EX537581
ACCESSION EX537581.1 GI:31873659
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1740)
Folb, G., Han, M., and Wiemann, S.
Submitted (17-JUN-2003) MRP, Ingolstaedter Landstr.1, D-85764
Neuberg, GERMANY
Neuberg, GERMANY
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
sequenced by GBF (National Research Centre for Biotechnology Ltd.,
Braunschweig/Germany) within the CDNA sequencing consortium of the
German Genome Project.
This clone (DKFZp686C04213) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at <http://mips.gsf.de/proj/cdna/>.

FEATURES

source 1..1740
/organism="Homo sapiens"
/mol_type="mRNA"

```

/db_xref="taxon:9606"
/clone="BRFZp686C04213"
/tissue_type="human rectum tumor"
/clone_id="686 (synonym: hicc3). Vector pSport1_Sfi; host
DH10B; sites SfiIa + SfiIb"
/dev stage="adult"
1694 .1699
polya_signal
polya_site 1720

```

```

Query Match 71.4%; Score 1029; DB 9; Length 1740;
Best Local Similarity 98.6%; Pired. No. 4.5e-207;
Matches 1038; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

```

```

QY 381 CTTCGCGGGGATCGGCGCACTGCGGCTGTTGGTGGTGAAGATTCTTCAGTCT 440
Db 688 CTGACCCGGGTGCTTCTTCGACGCGCTGGTTTGGTGGTGAAGATTCTTCAGTCT 747
QY 441 GTTCATAGGCGCAGAAATTGTGGCTGTGCACTTCAGTGAAGATTGTGGTGAAGT 500
Db 748 GTTCATAGGCGCAGAAATTGTGGCTGTGCACTTCAGTGAAGATTGTGGTGAAGT 807
QY 501 GAACACCAACACATCCCTCAACAAGCTTCAGCGAGCGCGGTTCACGCCCGTGGTCT 560
Db 808 GAACACCAACACATCCCTCAACAAGCTTCAGCGAGCGCGGTTCACGCCCGTGGTCT 867
QY 561 GCTCGTGGGCTGAGAGGCACTTAATATTACCTACACAGAGGCCCATCAGTGA 620
Db 868 GCTCGTGGGCTGAGAGGCACTTAATATTACCTACACAGAGGCCCATCAGTGA 927
QY 621 CGAACCATTGACTACACAGAGGCTTCACTGGCGCTTGAAGAATTTAGCGCGGA 680
Db 928 CGAACCATTGACTACACAGAGGCTTCACTGGCGCTTGAAGAATTTAGCGCGGA 987
QY 681 GTACGCGAAGCACTGAGAAAGGGGCTGCGGACCCAGTGTCTACCTGGCGAGAGTT 740
Db 988 GTACGCGAAGCACTGAGAAAGGGGCTGCGGACCCAGTGTCTACCTGGCGAGAGTT 1047
QY 741 CACACCGAGTACCCCTTGGCGCTGTACACACAGTACCTGGCGGACACTACGCTTC 800
Db 1048 CACACCGAGTACCCCTTGGCGCTGTACACACAGTACCTGGCGGACACTACGCTTC 1107
QY 801 GGCCACGCTATGGGTGGGCTTCTGCTTCTGGCTCTCTCCAGCTGTGCTCCAGGCT 860
Db 1108 GGCCACGCTATGGGTGGGCTTCTGCTTCTGGCTCTCTCCAGCTGTGCTCCAGGCT 1167
QY 861 GGCCCGCTCTAGCGAGGCTTGGCACTGTGACCAACCGAGCCCTTGGCGCTTCGAGGCT 920
Db 1168 GGCCCGCTCTAGCGAGGCTTGGCACTGTGACCAACCGAGCCCTTGGCGCTTCGAGGCT 1227
QY 921 CTTCGCGCTTGGCTTCATCTCTACAGTGGCGCTTGGCGCTGCGCTAGGCTCTCCGCT 980
Db 1228 CTTCGCGCTTGGCTTCATCTCTACAGTGGCGCTTGGCGCTGCGCTAGGCTCTCCGCT 1287
QY 981 GCTCAACAACAGTACGCGCGCGCTTCTTGGGTACGCTTGGCAACCGCGCTCTGTGCTT 1040
Db 1288 GCTCAACAACAGTACGCGCGCGCTTCTTGGGTACGCTTGGCAACCGCGCTCTGTGCTT 1347
QY 1041 CTTCCTGGAGGGGCGGTGAGTCTCCAGTATGTTCCGCCAGCGCTTGGCAACCGCT 1100
Db 1348 CTTCCTGGAGGGGCGGTGAGTCTCCAGTATGTTCCGCCAGCGCTTGGCAACCGCT 1407
QY 1101 TCTGACCAAAAGCGCCAGAGACTGAGCAGAGAGAGAGGGGCTCACTTTATCTCTCG 1160
Db 1408 TCTGACCAAAAGCGCCAGAGACTGAGCAGAGAGAGAGGGGCTCACTTTATCTCTCG 1467
QY 1161 CGAACCACTGACAGAGAGCGCGCTTCCAGACTTAAATATGATACCACTACTCTGG 1220
Db 1468 CGAACCACTGACAGAGAGCGCGCTTCCAGACTTAAATATGATACCACTACTCTGG 1527
QY 1221 AGGGGAGCCCAATCTGACTCTTCCCGCTTGGGACATGCGAGCGCGGAGCACTGC 1280

```

```

Db 1528 AGGGGAGCCCAATCTGACTCTTCCCGCTTGGGACATGCGAGCGCGGAGCACTGC 1587
QY 1281 CGCCGAGGCTGGGCGCAGAGAGCTCCAGAGAGGCACTGAGCGCTGTGGCGAGAGCC 1340
Db 1588 CGCCGAGGCTGGGCGCAGAGAGCTCCAGAGAGGCACTGAGCGCTGTGGCGAGAGCC 1647
QY 1341 TCGGACATCCGAGGAGCAGGAGAAAGTCTCTGGGGGAGATCTGTAATAAATCTTTT 1400
Db 1648 TCGGACATCCGAGGAGCAGGAGAAAGTCTCTGGGGGAGATCTGTAATAAATCTTTT 1707
QY 1401 TCTTTGTTTTTAAAAAATTTAAAAAATTTAAAAA 1433
Db 1708 TCTTTGTTTTTAAAAAATTTAAAAAATTTAAAAA 1740

```

```

RESULT 3
BC031111 1354 bp mRNA linear ROD 04-NOV-2003
LOCUS MGC:4989815, complete cds.
DEFINITION Mus musculus RIKEN cDNA 9030623N16 gene, mRNA (cDNA clone MGC:35715
ACCESSION BC031111 GI:21411407
VERSION BC031111
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

```

REFERENCE

1 (bases 1 to 1354)
Straussberg, R.L., Collins, F.S., Wagner, L., Sherman, C.M., Schuler, G.D., Klausner, R.D., Felling, B., Buecaw, K.H., Schaefer, C.F., Bhat, N.K., Altschul, S.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.F., Casavant, T.L., Scapleton, M., Soares, M.B., Bonaldi, M.F., Toshimichi, S., Schaefer, T.E., Brownstein, M.J., Usdin, T.B., Toshimichi, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McMan, P.J., McKernan, K.J., Malek, J.A., Guarniere, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hally, S.W., Vallalov, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fehey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, M.G., Rodriguez, A.C., Grimmond, J., Schmitt, J., Myers, R.M., Butcherfield, Y.S., Krzyzanski, M.I., Skalska, U., Small, D.E., Schercher, A., Schein, J.B., Jones, S.J., and Marra, M.A.

TITLE

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL

Medline
22388257
12477932

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
Contact: MGC help desk
Email: gcaps-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-ehgc.stanford.edu>
Contact: (Dickson, Mark) md@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>

Series: IRAX Plate: 58 Row: 9 Column: 17
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA 91: 13385245.
 Location/Qualifiers

FEATURES

source

1..1354

/organism="Mus musculus"

/mol_type="mRNA"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="MGC:35715 IMAGE:499815"

/tissue_type="Colon, normal, 5 month old male mouse."

/clone_id="NCI CGAP_C024"

/lab_host="DH10B"

/note="vector: pCMV-SPORT6"

1..1354

/gene="903062316R1X"

/db_xref="locusID:68811"

/db_xref="MGI:1914061"

129..1091

/codon_start=1

/product="RIKEN cDNA 9030623N16"

/protein_id="AAH3111.1"

/db_xref="GI:2411408"

/translation="MTANDVLPYPPQPRHASFVPLIVLVFSLAASPIPIIPG
 IRGSRMFLVAVLLEIFGAEIVAVHPSGDMFVGRVMTSYKAFSPSVQVVLH
 VGLAVNITLRTGPROQNETIDYNERFTWLNEDYVHLEKGLPDIYALEK
 FPPSPGCHGYOHLAGHAAATLVAECFWIINALSLMPEYGGALLTGGATL
 PGVFAFASISVPLCHPRLGSAVLTPPYGAFMLTLANGILSLIGAVVILHYTPS
 ALRSLFLDSVKDCNSQAKNSPILTNQHQLSPLNITTL"

ORIGIN

Query Match 49.4%; Score 711.4; DB 10; Length 1354;
 Best Local Similarity 75.9%; Pred. No. 6.6e-140; Indels 16; Gaps 4;
 Matches 936; Conservative 0; Mismatches 281;

222 CTGCCGCTGCTGAGACATCCGCGCGCGCTGACGATGACCTGCGGAAAGGCGACT 281
 92 CTGCGCTGCTGAGAAAGTTCTGAACTTTGACGATGACTGCGGAGGAGGCT 151
 282 GCGTTTAAACCCAGCCCGGAGATGCGGAGCTTCAAGCTTCACTCATCTTAT 341
 152 ACCCTTTAACCAGCCAGGATGCGGAGCTTCAAGCTTCACTCATCTTAT 211
 342 TCTAGTGTGCTGCTGAGACGAGCTTCTGCTCATCTTCCGGGAGATCCGTGCA 401
 212 CTGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 271
 402 CTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 461
 272 CTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 331
 462 GCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 521
 332 GCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 391
 522 AGCGTTCAGCGAGCGCGCTTACAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCT 581
 392 AGCGTTCAGCGAGCGCGCTTACAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCT 451
 582 TAATATTACTACACAGGAGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 641
 452 TAACATTACACTCCGAGGAGACCCAGGAGCGAGCTGAGACGACCATTTGATTAATGA 511
 642 GCAATTCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 701
 512 GCGTTCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 571
 702 GGGGCTGCGGAGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 761
 572 GGGAGCTGCGGAGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 631
 762 CTTGTACACCAAGTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 821

632 GCTTACACCAATATATACCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 691
 822 CTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 881
 692 CTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 751
 882 GGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 941
 752 GGCCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 811
 942 TAGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1001
 812 CAGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 871
 1002 GCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1061
 872 CTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 931
 1062 GAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1121
 932 GATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 991
 1122 CT---GCGAGGAG 1178
 992 CTGAGCAACAG 1051
 1179 GAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1238
 1052 GTTGAAGAGTCCAGACTTAA---TATTACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1108
 1239 CTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1297
 1109 TTTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1168
 1298 GAGAGAGTCCAG 1357
 1169 GATAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1228
 1358 CC-----AGGAGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1408
 1229 CTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1288
 1409 TTTTAAAG 1441
 1289 TTTTCTTCAAG 1321

RESULT 4
 AF547266 5851 bp mRNA linear MAM 03-APR-2003
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Sus scrofa (pig)
 Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
 1 (bases 1 to 5851)
 Morand, S., Charazou, M., Kanlewska, J., Deme, D., Ohayon, R.,
 Noel-Hudson, M.-S., Vairon, A. and Dupuy, C.
 Effect of iodine on nicotinamide adenine dinucleotide phosphate
 oxidase activity and Dnax2 protein expression in isolated porcine
 thyroid follicles
 Endocrinology 144 (4), 1241-1248 (2003)
 JOURNAL
 MEDLINE
 PUBMED
 2 (bases 1 to 5851)
 Kanlewska, J., Morand, S., Noel-Hudson, M.-S., Ohayon, R., Vairon, A.
 and Dupuy, C.
 Direct Submision
 TITLE
 JOURNAL
 Submitted (13-SEP-2002) Unite 486, INSERM University of Paris-Sud,

FEATURES

Location/Qualifiers

Source

1..5851

/organism="Sus scrofa"

/mol_type="mRNA"

/db_xref="taxon:9823"

1..5851

/gene="Duox1"

948..5609

/gene="Duox1"

/EC_number="1.6.3"

/function="putatively generates H2O2 by transferring electrons from NADH/NADPH to molecular oxygen"

/function="oxidoreductase acting on NAD(P)H with oxygen as acceptor"

/note="NADPH:O2 oxidoreductase"

/codon_start=1

/product="dual oxidase 1"

/protein_id="AA09338.1"

/db_xref="GI:23664373"

/translation="MGRFLALAWTLVGPMPMGANNSISWEVQRPDGMNMLMEKMGSKSRICQLVPASVADGVQPIGEPLHPYPRDLSTYMARAGAGASLAPRYVLGVPFGYHVLDSLEKRGCPAEFLNTHIPGDPYDPKSGSVYTPFGSRDPYTGSPSPNPRDLNTEVGMIDGSAIVGSSHSWSDERLFSFGQGLASGPAPRQMDLEFWTPDPATGQRGQGLAFGAEQGNREPLQALGLMFRYHNLCQKARERHLDGDELFQHAKRYIATVQSIIMYEMLPFLRKMPQBYTGYRFLDPSISPEFLASBQFSTWPGVYKMSACHFGQVIRNNSVSRALRVNCSYMSREHNLORAEVDALIGMSQIAERDHWVVDVDFPWRGPIKFSRTHDASCLORGDLHSYTRARALGIPVTRMODITPALSRSDDGLYELTALYNQDLSRELLFGGLSEYSGPGLFSTIVDQVRLRQDRVYFENTKNGLFSEKEIAIRNLSRSDVLAATNTFPALQDNVFWAGDPCQPRQLSTKDLPAKAPILMDYFKSGSGFVLTGLCPFLVLSLAAVAQLRNFKRLQVQNRQSLMEKLVGMKALEMQGRKPCRPVLSQGIHMDRLSVLTLQLRPQVNLISLMSHGRRLILKIPEYDLVLFDEERQVWVNTQALSESLISQCEMLRBOELMARAVTRESORSHLETFRLFSQVLDIQAGALPLDSQKVEALTCELSRAPASLGIKPDQMFESMSLAKDNGYLSREFDLIVPMKSPREKSRLMRMTDPDNGELISDEIRIMLRSTIEISNCSKAQUTEVSEMPREAGPQKORLIMEDPFMLRDHSELFTQLCVKGVVEVATKDCRASIYSGEKLCPSRYNACPRSNVDEVELTPWKLQCPDTPDPOIRRRPGRKVTYSQPLIFEAHERKQSRHQTVOQFRFVENYRHHGLAVFYIAGLFLERAVYVFAAHMILDTTGVILISRGTAASISFMSFYLITMCENLITFLAETPLNRYVFDAAVPHRIJASTALITVLHSAGVNVVYLSISPLSVSLCPFLGHONGSEFPQKYVMFPQVTPGVLMLLILAIWYFASHHRRCSFRGFWLTHLVLYLMLLIHGSFPLDLPFRHFFVLPALITYGDKLVLISRKKEISVYKALPLPSGTHQQRPRPGREFKSGQWRICACLGTTXYHPFTLSAPHEDLTSLHRAAGPWTLRITSPYDDNCARPKLYDGPGECHQEHWHKEFVSLVGGGIGVTFPASILKDLVFKSSVSCQVCKKIYFIWTRQRPFWLADIIREVENDDHDLVSAHYITOLAEKFDLRTMLVYCEHFGKVNRSFTGLRSITHFGRPPFPFPFNSLOEVHPQVKIGVFCGDPGMTXKVERACQLINQDRTHFSHYENF"

ORIGIN

Query Match 43.5%; Score 626.2; DB 4; Length 5851;
Best Local Similarity 81.8%; Pred. No. 7.6e-122;
Matches 735; Conservative 0; Mismatches 163; Indels 1; Gaps 1;

254 GCAGATGACCGCTGGAGAACGGCGACGCTTTTACCCCGACGCCGCGCATGCCGAG 313
Db GCAGATGACCTGTGGAGAACGGCGTGCCTTCTACCTCAGCCTCGGTAGCCGCCA 69
314 GCTTACGCGTTCCATCGTCTATCTAGTGTGTTGGCTTAGCAGCAAGCTTCC 373
Db GCTTACGCGTTCCATCGTCTATCTAGTGTGTTGGCTTAGCAGCAAGCTTCC 129
374 TGCTGATCTTGGCCGGGATCGGTGGCACTCGCGCTGTTTGGTGGTAGAGTCTTC 433
Db TACTGATCTTGGCCGGGATCGGTGGCACTCGCGCTGTTTGGTGGTAGAGTCTTC 189
434 TCAGTCTGTATAGCGCCAGAAATTGGCTGTGCACTTCACTGACGAATGGTGTG 493
Db TAGCTGTGTATAGCGCCAGAAATTGGCTGTGCACTTCACTGACGAATGGTGTG 249
494 GTACAGTGAACCAACATCTTCAAAAGCTTCAGCGAGCGCGGTACAGCCCGTG 553
Db GCAAGTTAGCAACATCTTCAAAAGCTTCAGCGAGCGGTACAGCCCGTG 309

554 TCGCTCTGCTCGTGGGCTTGAAGGCATTAAATTACACTCAGAGGACCCCAAGTGC 613
Db TCGCTCTGACCGGGCTTGAAGGCTTGAATATACACTCAGAGGAAATCCGATACG 369
614 AGCTGACGAGACCATTTGACTACCAAGCAGATTCAGCTGGCTTGAAGGAAATTACG 673
Db AGCTGAATGAGACCATTCAGACTCAAGAGCAATTCATCTGGGATATGGGCGAAACTATG 429
674 CCGGAGATACGCGAACCGCATGAGAGAGAGGCTCCGAGACCAATGCTTACCTGGCGG 733
Db CCGGAGGATATGGAAGGCTTGAAGAGAGGCTCCGAGACCAATGCTTACCTGGCGG 489
734 AGAAGTTACACAGGATAGCCCTTGGGCGCTTACACAGCATACCACTGGCGGAGACT 793
Db AGAAGTT-ATCCAGAAAGCCCTTGGGCGCTTACCGCAATATGCGCTAGCGGAGACT 548
794 ACGCTTCGCGACGCTATGAGGTGGCTTCTGCTTCTGCTCTTCCAAAGTGTGCTCT 853
Db ACGCTTCGCGACATTTGTGAGCGGCTTCTGCTTCTGACTCTTCCAAAGTGTGCTCT 608
854 CCAAGCGCGCGCGCTCTAGAGAGGCTGAGCACTGCTGACCAAGAGGCTTGGCGCTCT 913
Db CCAATGCGGCTCCGACATTTGTGAGCGGCTTCTGCTTCTCTATCACCGGCGCTTCT 668
914 TCGGAGTCTTGGCTTGGCTTCCATCTCTAGCGGTGCGCTTGGCGCTTGGCGCTT 973
Db TCTGTGCTTGGCTTGGCTTCTCTATCTCAGCGTGTCTTGGCAGCTTGGCGCTT 728
974 CTTCCGCGCTCAGCTCAGTACAGGCGCGCTTCTGGGTACAGGTGGAACCGCGCT 1033
Db CTTCCGAGATCAGCTCAGTACAGGCGCGCTTCTGGGTACAGGTGGAACCGCGCT 788
729 CTTCCGAGATCAGCTCAGTACAGGCGCGCTTCTGGGTACAGGTGGAACCGCGCT 788
Db TGTGCTCTCTCTCGGAGAGCGGTGTGATGCTCAGTACAGCTCAGCTCAGCT 848
1034 TGTGCTCTCTCTCGGAGAGCGGTGTGATGCTCAGTACAGGTGGAACCGCGCT 1093
Qy 1094 GCACCTCTTGAACCAAGCGCGCAAGCATGCAAGCAGAGAGAGGCGCTCACTCTT 1152
Db 849 GCAGTCTTGAAGGAGATGATCAAGCATGCAAGCAGAGAGAGGCGCTCACTTT 907

RESULT 5

AX035348 498 bp DNA linear PAT 15-NOV-2000

LOCUS AX035348

DEFINITION Sequence 3 from Patent WO053748.

ACCESSION AX035348

VERSION AX035348.1 GI:11191065

KEYWORDS

SOURCE

ORGANISM Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Bruck,C.E., Vinals,Y.D., Coche,T. and Caesart,J.P.

TITLE Novel compounds

JOURNAL Patent: WO 0053748-A 3 14-SEP-2000;

BRUCK CLAUDEINE ETVIRE MARIE (BE) ; SMITHLINE BECHAM BIOLOG (BE) ; VINALS Y DE BASSOLS CARLOTTA (BE) ; COCHE THIERRY (BE) ; CASBART JEAN POL (BE)

FEATURES

source

1..498

/organism="Homo sapiens"

/mol_type="unassigned DNA"

/db_xref="taxon:9606"

ORIGIN

Query Match 32.3%; Score 465.4; DB 6; Length 498;
Best Local Similarity 99.4%; Pred. No. 6.6e-88;
Matches 488; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

939 CTTAGGCTGCGGCTTGGCGGCT-CGGCTAGGCTCTTGGCGGCTCAGCAGTACG 997

```

Db      1  CTTAGAGCTGCGGCTCTGCGCGCTCCGCGCTAGAGCTCTCCGCGCTCACCACTCAGTACG 60
QY      998 -CGCCGCGCTTTGGGATCAGCTGCGCAACCGCGCTCTGTGCTCTTCTTCTGAGAGGCGCC 1056
Db      61  AGCGCGCGCTTTGGGATCAGCTGCGCAACCGCGCTCTGTGCTCTTCTTCTGAGAGGCGCC 120
QY      1057 GGGATGAGTCTCAGATATTTGGGCGCGCTCTTGGCGACCTTTCTGACCAAGAGCGCC 1116
Db      121  GGGATGAGTCTCAGATATTTGGGCGCGCTCTTGGCGACCTTTCTGACCAAGAGCGCC 180
QY      1117 AAGAGCTGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1176
Db      181  AAGAGCTGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
QY      1177 CAGGCGCGCTCTCCGAGACTTAAAGTATCACCCTTATCTTCTGAGAGAGAGAGAGAGAGAG 1236
Db      241  CAGGCGCGCTCTCCGAGACTTAAAGTATCACCCTTATCTTCTGAGAGAGAGAGAGAGAGAG 300
QY      1237 GACTCTCTCCGCGCTTTGGGACATCGACAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1296
Db      301  GACTCTCTCCGCGCTTTGGGACATCGACAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
QY      1297 AAGAGAGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1356
Db      361  AAGAGAGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
QY      1357 ACCAGGAGAAAGTCTCTGCGGCGCATCTGTAATAAACCTTTTCTTTCTTTTCTTTTAA 1416
Db      421  ACCAGGAGAAAGTCTCTGCGGCGCATCTGTAATAAACCTTTTCTTTCTTTTCTTTTAA 480
QY      1417 AAAAAAAAAA 1427
Db      481  AAAAAAAAAA 491

RESULT 6
AC012255 166937 bp DNA linear HTG 06-MAY-2001
LOCUS     Homo sapiens chromosome 15 clone RP11-109D20 map 15, WORKING DRAFT
DEFINITION
SEQUENCE 22 unordered pieces.
AC012255 4 GI:7684440
VERSION   HTG: HTGS PHASE1; HTGS_DRAFT.
KEYWORDS  Homo sapiens (human)
SOURCE    Homo sapiens
ORGANISM  Homo sapiens
REFERENCE
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
TITLE     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL   Unpublished
AUTHORS   2 (bases 1 to 166937)
          1 (bases 1 to 166937)
          Homo sapiens chromosome 15, clone RP11-109D20
          Unpublished
          2 (bases 1 to 166937)
          Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
          Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Bokhgalter,B.,
          Brown,A., Castle,A., Colangelo,M., Collins,S., Collins,A.,
          Cooke,P., Dearellano,K., Dewar,K., Domingo,M., Donnelly,M., Doyle,M.,
          Farrell,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
          Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heathford,A., Horton,L.,
          Howland,J.C., Johnson,R., Jones,C., Kann,L., Karasch,A., Klein,J.,
          Lehotsky,J., Lien,C., Locke,K., MacDonald,P., Margulis,N.,
          McMan,P., McGurk,A., McKernan,K., McLaughlin,D., Meltrin,J.,
          Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
          Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
          Stenge-Thomann,N., Stojanovic,N., Subramanian,A., Talmage,J.,
          Telford,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
          Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
COMMENT   Direct Submission
          Submitted (21-OCT-1999) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
          On May 3, 2000 this sequence version replaced gi:7209328.
          All repeats were identified using RepeatMasker:
          Smit, A.F.A. & Green, P. (1996-1997)
          http://ftp.genome.washington.edu/RV/RepeatMasker.html

```

```

FEATURES
source
* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 2087: contig of 2087 bp in length
2088 2187: gap of 100 bp
2188 3266: contig of 1439 bp in length
3267 3726: gap of 100 bp
3727 4845: contig of 1119 bp in length
4846 4946: gap of 100 bp
4946 6702: contig of 1757 bp in length
6703 6803: gap of 100 bp
6803 8327: contig of 1524 bp in length
8327 8427: gap of 100 bp
8427 10494: contig of 2068 bp in length
10495 10594: gap of 100 bp
10595 13215: contig of 2621 bp in length
13216 13315: gap of 100 bp
13315 16451: contig of 3136 bp in length
16451 16551: gap of 100 bp
16551 20843: contig of 4292 bp in length
20844 20943: gap of 100 bp
20943 24940: contig of 3997 bp in length
24941 25040: gap of 100 bp
25040 25841: contig of 4861 bp in length
25841 29901: gap of 100 bp
29902 30002: gap of 100 bp
30002 34189: contig of 4188 bp in length
34190 34289: gap of 100 bp
34289 38386: contig of 4097 bp in length
38387 38486: gap of 100 bp
38486 45827: contig of 7341 bp in length
45828 45927: gap of 100 bp
45928 51729: contig of 5802 bp in length
51730 51829: gap of 100 bp
51829 60123: contig of 8294 bp in length
60124 60223: gap of 100 bp
60223 73532: contig of 13309 bp in length
73533 73632: gap of 100 bp
73632 87433: contig of 13801 bp in length
87434 87533: gap of 100 bp
87533 101708: contig of 14175 bp in length
101709 101808: gap of 100 bp
101809 121193: contig of 19385 bp in length
121194 121293: gap of 100 bp
121294 139051: contig of 17758 bp in length
139052 139151: gap of 100 bp
139152 166937: contig of 27786 bp in length.
Location/Qualifiers
1.166937
/organism="Homo sapiens"
/mol_type="genomic DNA"

```

```

/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-109D20"
/clone_lib="RPC1-11 Human Male BAC"
1..2087
/note="assembly_fragment"
misc_feature
2188..3626
/note="assembly_fragment"
3727..4845
/note="assembly_fragment"
4946..6702
/note="assembly_fragment"
6803..8326
/note="assembly_fragment"
8427..10494
/note="assembly_fragment"
10595..13215
/note="assembly_fragment"
13316..16451
/note="assembly_fragment"
16552..20843
/note="assembly_fragment"
20944..24940
/note="assembly_fragment"
25041..29901
/note="assembly_fragment"
30002..34189
/note="assembly_fragment"
34290..38386
/note="assembly_fragment"
38487..45827
/note="assembly_fragment"
45928..51729
/note="assembly_fragment"
51830..60123
/note="assembly_fragment"
60224..73532
/note="assembly_fragment"
73633..87433
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
87534..101708
/note="assembly_fragment"
101809..121193
/note="assembly_fragment"
121294..139051
/note="assembly_fragment"
139152..166937
/note="assembly_fragment"
clone_end:T7
vector_side:right"

ORIGIN
Query Match      28.1%; Score 405.4; DB 2; Length 166937;
Best Local Similarity 98.6%; Pred. No. 4.7e-75;
Matches 409; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY      1 AAAGTAA CGGCTA CAGACA GTGA GAATTA GTTGCT GCCTGC CGCGCT AGAAAA CTCTGT CG 60
Db      157716 AAAGTAA CGGCTA CAGACA GTGA GAATTA GTTGCT GCCTGC CGCGCT AGAAAA CTCTGT CG 157715
QY      61 GTACCA ACCCAGA GCGTTA GAGAGC GCCCAC CTCACAG CTTGCC TTAACG GAGAGG TGCA 120
Db      157776 GTACCA ACCCAGA GCGTTA GAGAGC GCCCAC CTCACAG CTTGCC TTAACG GAGAGG TGCA 157835
QY      121 GGACTCA GACTTC ACGACG CCATCC GATCCC ACCTTGT ATCGCA AAGAGA CGCCAG CAAGAC 180
Db      157836 GGACTCA GACTTC ACGACG CCATCC GATCCC ACCTTGT ATCGCA AAGAGA CGCCAG CAAGAC 157899
QY      181 GGCGTCT CCCGGCT CCAAGG AGCCCC AGCTTGT GCTG ACTTGC CTGCTG CCGCCT GCGCTG CAGC 240

```

Db	157896	GGGCTCTCCCGGTCGACAGACGCCACAGCTTGAGCTGGCTGTCGCCGCTGGGTGAG	1579955		
Oy	241	ACTCGGCGGGCTGTCAGCATACCTGTGTGGAACGGCGTACTGCTTTTACCCCAAGCC	300		
Db	157956	ACTCGGCGGGCTGTCAGCATGACCTGTGGAACGGCGTACTGCTTTTACCCCAAGCC	158015		
Oy	301	CGGCAATCGCGAGGCTTCAGCTTCATGCTCATTCGTATTCAAGTGTTTTGGCTCTA	360		
Db	158016	CGGCAATCGCGAGGCTTCAGCTTCATGCTCATTCGTATTCAAGTGTTTTGGCTCTA	1580755		
Oy	361	GCAGCAAGCTTCTGCTCATCTTGGCCGGGATCCGTGGCCACTGGCGCTGTTT	415		
Db	158076	GCAGCAAGCTTCTGCTCATCTTGGCCGGGATCCGTGGCCACTGGGTAGGGTGT	158130		
RESULT 7					
AC091117		181312 bp	DNA	linear	PRI 20-FEB-2002
LOCUS					
DEFINITION					Homo sapiens chromosome 15 clone RP11-109D20 map 15q15, complete sequence.
AC091117					
VERSION					AC091117.5
KEYWORDS					GI:18767395
SOURCE					HMG.
ORGANISM					Homo sapiens (human)
REFERENCE					
AUTHORS					Homio sapiens
TITLE					Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
JOURNAL					Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE					1 (bases 1 to 181312)
AUTHORS					Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
TITLE					Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
JOURNAL					Pate,D., and Hood,L.
REFERENCE					Sequencing of human chromosome 15 D15S146-D15S117 region
AUTHORS					Unpublished
TITLE					2 (bases 1 to 181312)
JOURNAL					Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
REFERENCE					Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
AUTHORS					Pate,D., and Hood,L.
TITLE					Direct Submision
JOURNAL					Submitted (13-FEB-2002) Multimegabase Sequencing Center, Institute
REFERENCE					for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
AUTHORS					98105, USA
TITLE					3 (bases 1 to 181312)
JOURNAL					Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
REFERENCE					Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
AUTHORS					Pate,D., and Hood,L.
TITLE					Direct Submision
JOURNAL					Submitted (14-FEB-2002) Multimegabase Sequencing Center, Institute
REFERENCE					for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
AUTHORS					98105, USA
TITLE					4 (bases 1 to 181312)
JOURNAL					Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
REFERENCE					Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
AUTHORS					Pate,D., and Hood,L.
TITLE					Direct Submision
JOURNAL					Submitted (14-FEB-2002) Multimegabase Sequencing Center, Institute
REFERENCE					for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
AUTHORS					98105, USA
TITLE					5 (bases 1 to 181312)
JOURNAL					Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
REFERENCE					Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
AUTHORS					Pate,D., and Hood,L.
TITLE					Direct Submision
JOURNAL					Submitted (20-FEB-2002) Multimegabase Sequencing Center, Institute
REFERENCE					for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
AUTHORS					98105, USA
TITLE					On Feb 20, 2002 this sequence version replaced gi.18653547.
JOURNAL					
COMMENT					
					----- Genome Center
					Center: Multimegabase Sequencing Center
					Center code: UMSC
					Web site: http://chroma.mbt.washington.edu/msg_www
					Contact: leeirowen@systemsbiology.org
					----- Summary Statistics
					Sequencing vector: pUC18; U08752

***** NOTE: This is a 'working draft' sequence. It currently consists of 15 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 1169: contig of 1169 bp in length
1170 1269: gap of 100 bp
1270 3335: contig of 2066 bp in length
3336 3435: gap of 100 bp
3436 5698: contig of 2263 bp in length
5699 5798: gap of 100 bp
5799 9215: contig of 3418 bp in length
9217 9315: gap of 100 bp
9317 15633: contig of 6317 bp in length
15634 15733: gap of 100 bp
15734 21746: contig of 6013 bp in length
21747 21846: gap of 100 bp
21847 30351: contig of 8505 bp in length
30352 30452: gap of 100 bp
30453 36509: contig of 6058 bp in length
36510 36609: gap of 100 bp
36610 45279: contig of 8670 bp in length
45280 45379: gap of 100 bp
45380 56851: contig of 11472 bp in length
56852 67375: gap of 100 bp
67376 67377: contig of 10423 bp in length
67378 79534: gap of 12060 bp in length
79535 79634: gap of 100 bp
79635 95432: contig of 15798 bp in length
95433 95532: gap of 100 bp
95533 120801: contig of 25269 bp in length
120802 120901: gap of 100 bp
120902 156534: contig of 35633 bp in length.
Location/Qualifiers
1. 156534
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-163P10"
/clone_id="RPC1-11 Human Male BAC"
1. 1169
/feature="assembly_fragment"
1270. 3335
/feature="assembly_fragment"
3436. 5698
/feature="assembly_fragment"
5799. 9216
/feature="assembly_fragment"
9317. 15633
/feature="assembly_fragment"
15734. 21746
/feature="assembly_fragment"
21847. 30351
/feature="assembly_fragment"
30452. 36509
/feature="assembly_fragment"
36610. 45279
/feature="assembly_fragment"
45280. 45379
/feature="assembly_fragment"
45380. 56851
/feature="assembly_fragment"
56852. 67375
/feature="assembly_fragment"
67376. 79534
/feature="assembly_fragment"
79535. 95432
/feature="assembly_fragment"
95433. 95532
/feature="assembly_fragment"
95533. 120801
/feature="assembly_fragment"
120802. 120901
/feature="assembly_fragment"
120902. 156534
/feature="assembly_fragment"

```

```

misc_feature 79635..95432
/feature="assembly_fragment"
misc_feature 95533..120801
/feature="assembly_fragment"
misc_feature 120902..156534
/feature="assembly_fragment"
misc_feature 156535..163534
/feature="assembly_fragment"
misc_feature 163535..174444
/feature="assembly_fragment"
misc_feature 174445..185555
/feature="assembly_fragment"
misc_feature 185556..196666
/feature="assembly_fragment"
misc_feature 196667..207777
/feature="assembly_fragment"
misc_feature 207778..218888
/feature="assembly_fragment"
misc_feature 218889..229999
/feature="assembly_fragment"
misc_feature 229999..241111
/feature="assembly_fragment"
misc_feature 241112..252222
/feature="assembly_fragment"
misc_feature 252223..263333
/feature="assembly_fragment"
misc_feature 263334..274444
/feature="assembly_fragment"
misc_feature 274445..285555
/feature="assembly_fragment"
misc_feature 285556..296666
/feature="assembly_fragment"
misc_feature 296667..307777
/feature="assembly_fragment"
misc_feature 307778..318888
/feature="assembly_fragment"
misc_feature 318889..329999
/feature="assembly_fragment"
misc_feature 329999..341111
/feature="assembly_fragment"
misc_feature 341112..352222
/feature="assembly_fragment"
misc_feature 352223..363333
/feature="assembly_fragment"
misc_feature 363334..374444
/feature="assembly_fragment"
misc_feature 374445..385555
/feature="assembly_fragment"
misc_feature 385556..396666
/feature="assembly_fragment"
misc_feature 396667..407777
/feature="assembly_fragment"
misc_feature 407778..418888
/feature="assembly_fragment"
misc_feature 418889..429999
/feature="assembly_fragment"
misc_feature 429999..441111
/feature="assembly_fragment"
misc_feature 441112..452222
/feature="assembly_fragment"
misc_feature 452223..463333
/feature="assembly_fragment"
misc_feature 463334..474444
/feature="assembly_fragment"
misc_feature 474445..485555
/feature="assembly_fragment"
misc_feature 485556..496666
/feature="assembly_fragment"
misc_feature 496667..507777
/feature="assembly_fragment"
misc_feature 507778..518888
/feature="assembly_fragment"
misc_feature 518889..529999
/feature="assembly_fragment"
misc_feature 529999..541111
/feature="assembly_fragment"
misc_feature 541112..552222
/feature="assembly_fragment"
misc_feature 552223..563333
/feature="assembly_fragment"
misc_feature 563334..574444
/feature="assembly_fragment"
misc_feature 574445..585555
/feature="assembly_fragment"
misc_feature 585556..596666
/feature="assembly_fragment"
misc_feature 596667..607777
/feature="assembly_fragment"
misc_feature 607778..618888
/feature="assembly_fragment"
misc_feature 618889..629999
/feature="assembly_fragment"
misc_feature 629999..641111
/feature="assembly_fragment"
misc_feature 641112..652222
/feature="assembly_fragment"
misc_feature 652223..663333
/feature="assembly_fragment"
misc_feature 663334..674444
/feature="assembly_fragment"
misc_feature 674445..685555
/feature="assembly_fragment"
misc_feature 685556..696666
/feature="assembly_fragment"
misc_feature 696667..707777
/feature="assembly_fragment"
misc_feature 707778..718888
/feature="assembly_fragment"
misc_feature 718889..729999
/feature="assembly_fragment"
misc_feature 729999..741111
/feature="assembly_fragment"
misc_feature 741112..752222
/feature="assembly_fragment"
misc_feature 752223..763333
/feature="assembly_fragment"
misc_feature 763334..774444
/feature="assembly_fragment"
misc_feature 774445..785555
/feature="assembly_fragment"
misc_feature 785556..796666
/feature="assembly_fragment"
misc_feature 796667..807777
/feature="assembly_fragment"
misc_feature 807778..818888
/feature="assembly_fragment"
misc_feature 818889..829999
/feature="assembly_fragment"
misc_feature 829999..841111
/feature="assembly_fragment"
misc_feature 841112..852222
/feature="assembly_fragment"
misc_feature 852223..863333
/feature="assembly_fragment"
misc_feature 863334..874444
/feature="assembly_fragment"
misc_feature 874445..885555
/feature="assembly_fragment"
misc_feature 885556..896666
/feature="assembly_fragment"
misc_feature 896667..907777
/feature="assembly_fragment"
misc_feature 907778..918888
/feature="assembly_fragment"
misc_feature 918889..929999
/feature="assembly_fragment"
misc_feature 929999..941111
/feature="assembly_fragment"
misc_feature 941112..952222
/feature="assembly_fragment"
misc_feature 952223..963333
/feature="assembly_fragment"
misc_feature 963334..974444
/feature="assembly_fragment"
misc_feature 974445..985555
/feature="assembly_fragment"
misc_feature 985556..996666
/feature="assembly_fragment"
misc_feature 996667..1007777
/feature="assembly_fragment"

```

```

misc_feature 79635..95432
/feature="assembly_fragment"
misc_feature 95533..120801
/feature="assembly_fragment"
misc_feature 120902..156534
/feature="assembly_fragment"
misc_feature 156535..163534
/feature="assembly_fragment"
misc_feature 163535..174444
/feature="assembly_fragment"
misc_feature 174445..185555
/feature="assembly_fragment"
misc_feature 185556..196666
/feature="assembly_fragment"
misc_feature 196667..207777
/feature="assembly_fragment"
misc_feature 207778..218888
/feature="assembly_fragment"
misc_feature 218889..229999
/feature="assembly_fragment"
misc_feature 229999..241111
/feature="assembly_fragment"
misc_feature 241112..252222
/feature="assembly_fragment"
misc_feature 252223..263333
/feature="assembly_fragment"
misc_feature 263334..274444
/feature="assembly_fragment"
misc_feature 274445..285555
/feature="assembly_fragment"
misc_feature 285556..296666
/feature="assembly_fragment"
misc_feature 296667..307777
/feature="assembly_fragment"
misc_feature 307778..318888
/feature="assembly_fragment"
misc_feature 318889..329999
/feature="assembly_fragment"
misc_feature 329999..341111
/feature="assembly_fragment"
misc_feature 341112..352222
/feature="assembly_fragment"
misc_feature 352223..363333
/feature="assembly_fragment"
misc_feature 363334..374444
/feature="assembly_fragment"
misc_feature 374445..385555
/feature="assembly_fragment"
misc_feature 385556..396666
/feature="assembly_fragment"
misc_feature 396667..407777
/feature="assembly_fragment"
misc_feature 407778..418888
/feature="assembly_fragment"
misc_feature 418889..429999
/feature="assembly_fragment"
misc_feature 429999..441111
/feature="assembly_fragment"
misc_feature 441112..452222
/feature="assembly_fragment"
misc_feature 452223..463333
/feature="assembly_fragment"
misc_feature 463334..474444
/feature="assembly_fragment"
misc_feature 474445..485555
/feature="assembly_fragment"
misc_feature 485556..496666
/feature="assembly_fragment"
misc_feature 496667..507777
/feature="assembly_fragment"
misc_feature 507778..518888
/feature="assembly_fragment"
misc_feature 518889..529999
/feature="assembly_fragment"
misc_feature 529999..541111
/feature="assembly_fragment"
misc_feature 541112..552222
/feature="assembly_fragment"
misc_feature 552223..563333
/feature="assembly_fragment"
misc_feature 563334..574444
/feature="assembly_fragment"
misc_feature 574445..585555
/feature="assembly_fragment"
misc_feature 585556..596666
/feature="assembly_fragment"
misc_feature 596667..607777
/feature="assembly_fragment"
misc_feature 607778..618888
/feature="assembly_fragment"
misc_feature 618889..629999
/feature="assembly_fragment"
misc_feature 629999..641111
/feature="assembly_fragment"
misc_feature 641112..652222
/feature="assembly_fragment"
misc_feature 652223..663333
/feature="assembly_fragment"
misc_feature 663334..674444
/feature="assembly_fragment"
misc_feature 674445..685555
/feature="assembly_fragment"
misc_feature 685556..696666
/feature="assembly_fragment"
misc_feature 696667..707777
/feature="assembly_fragment"
misc_feature 707778..718888
/feature="assembly_fragment"
misc_feature 718889..729999
/feature="assembly_fragment"
misc_feature 729999..741111
/feature="assembly_fragment"
misc_feature 741112..752222
/feature="assembly_fragment"
misc_feature 752223..763333
/feature="assembly_fragment"
misc_feature 763334..774444
/feature="assembly_fragment"
misc_feature 774445..785555
/feature="assembly_fragment"
misc_feature 785556..796666
/feature="assembly_fragment"
misc_feature 796667..807777
/feature="assembly_fragment"
misc_feature 807778..818888
/feature="assembly_fragment"
misc_feature 818889..829999
/feature="assembly_fragment"
misc_feature 829999..841111
/feature="assembly_fragment"
misc_feature 841112..852222
/feature="assembly_fragment"
misc_feature 852223..863333
/feature="assembly_fragment"
misc_feature 863334..874444
/feature="assembly_fragment"
misc_feature 874445..885555
/feature="assembly_fragment"
misc_feature 885556..896666
/feature="assembly_fragment"
misc_feature 896667..907777
/feature="assembly_fragment"
misc_feature 907778..918888
/feature="assembly_fragment"
misc_feature 918889..929999
/feature="assembly_fragment"
misc_feature 929999..941111
/feature="assembly_fragment"
misc_feature 941112..952222
/feature="assembly_fragment"
misc_feature 952223..963333
/feature="assembly_fragment"
misc_feature 963334..974444
/feature="assembly_fragment"
misc_feature 974445..985555
/feature="assembly_fragment"
misc_feature 985556..996666
/feature="assembly_fragment"
misc_feature 996667..1007777
/feature="assembly_fragment"

```

```

RESULT 9
AC087790
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

```

Query Match 28.0%; Score 403.8; DB 2; Length 156534;
Best Local Similarity 98.3%; Pred. No. 16-74; 7; Indels 0; Gaps 0;
Matches 408; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

1 AAAGTACGGCTTACAGACAGTGAATATGTTGCTGCGCGCTGAGAAAATCTGTG 60
7148 AAAGTACGGCTTACAGACAGTGAATATGTTGCTGCGCGCTGAGAAAATCTGTG 7089

61 GTACCAACCCAGAGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
7088 GTACCAACCCAGAGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7029

121 GGAAGTACAGCTTACAG 180
7028 GGAAGTACAGCTTACAG 6969

181 GCGCTCTCCGCGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
6968 GCGCTCTCCGCGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6909

241 ACTGCGCGCGCTGAG 300
6908 ACTGCGCGCGCTGAG 6849

301 CGGATACCGCGAGCGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
6848 CGGATACCGCGAGCGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6789

361 GCAGCAAGCTTCTGCTCATCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 415
6788 GCAGCAAGCTTCTGCTCATCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6734

AC087790 171444 bp DNA linear HTG 09-MAY-2001
Homo sapiens chromosome 15 clone RP11-276K9 map 15, WORKING DRAFT
SEQUENCE 13 unordered pieces.
AC087790 GI:13357344
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 171444)
Barnes, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
Barnes, N., Bassett, V., Boguslavsky, L., Bokil, B., Brown, A.,
Cammarata, J., Campiano, A., Choe, Y., Colangelo, M., Collins, S.,
Collins, A., Cooke, P., DeRubeis, K., Dewar, K., Diaz, J., Diaz, J.,
Dodge, S., Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J.,
Gardner, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heath, A., Horton, L., Hume, W., Iliev, I., Johnson, R.,
Jones, C., Karas, A., Labrecque, K., Lamazares, R., Landers, T.,
Leibach, J., Levine, R., Liu, G., Maclean, C., MacDonald, P.,
Marquis, N., Matthews, C., McCarthy, M., McEwen, P., McKernan, K.,
McNeeley, R., Meldrum, J., Menes, L., Mihov, T., Miska, V.,
Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Punkhankar, P., Pierre, N., Pollara, V., Raymond, C., Retter, R.,
Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,

Query Match	28.0%	Score 03.8	DB 2	Length 171444
Best Local Similarity	98.3%	Pred. No. 1e-74		
Matches 408	Conservative 0	Mismatches 7	Indels 0	Gaps 0
QY	1	AAAGTAA	CGGCTACAGACAGT	GAGAAATAGTTTCGCTCGCCGCGCTAGAAAAAATCTGTG 60
Db	140413	AAAGTAA	CGGCTACAGACAGT	GAGAAATAGTTTCGCTCGCCGCGCTAGAAAAAATCTGTG 140472
QY	61	GTACCAACCC	CAGACGTTTAGAGACAGCCACCTCCACGCTTCTTAACGAGAGTGCA 120	
Db	140473	GTACCAACCC	CAGACGTTTAGAGAGAGCCACCTCCACGCTTCTTAACGAGAGTGCA 140532	
QY	121	GGAGTCAGACTT	CACAGAGCCCACTCGGTCCAGGCTTGTAAGCAAAAGAGAGCCCAAGAC 180	
Db	140533	GGAGTCAGACTT	CACAGAGCCCACTCGGTCCAGGCTTGTAAGCAAAAGAGAGCTCAAGAC 140592	
QY	181	GGCGCTCTCC	CGGCGTCAAGGACCCCACTTGTCGTTCGCTCGCCGCTGCTGAGC 240	
Db	140593	GGCGCTCTCC	CGGCGTCAAGGACCCCACTTGTCGTTCGCTCGCCGCTGCTGAGC 140652	
QY	241	ACTCGGCGGCGGT	GCAGCATACCTGTGGAAACGGCGTACGCTTTTACCCCAAGCC 300	
Db	140653	ACTCGGCGGCGGT	GCAGCATACCTGTGGAAACGGCGTACGCTTTTACCCCAAGCC 140712	
QY	301	CGGCAATGCGGAG	GCTTACGCTTCACTGCTCATCGTTATTCATAGTTTGGCTCTA 360	
Db	140713	CGGCAATGCGGAG	GCTTACGCTTCACTGCTCATCGTTATTCATAGTTTGGCTCTA 140772	
QY	361	GCAGCAAGCTT	CTCTGCTATCTTGGCCGGGGAATCGGTGCACTCGCGCTGTTT 415	
Db	140773	GCAGCAAGCTT	CTCTGCTATCTTGGCCGGGGAATCGGTGCACTCGGTGAGGTGT 140827	

RESULT 10

AX714511/c 2684 bp DNA linear PAT 15-APR-2003

LOCUS AX714511

DEFINITION Sequence 1195 from Patient EP1293569.

ACCESSION AX714511

VERSION AX714511.1 GI:29889464

KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS
1 Isogai, T., Sugiyama, T., Otsuki, T., Makamatsu, A., Sato, H., Ishii, S., Yamamoto, J. I., Isono, Y., Hito, Y., Otsuka, K., Nagai, K., Irie, R., Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahara, K. and Masuko, Y.
TITLE Full-length cDNAs
JOURNAL Patent: EP 1293569-A 1195 19-MAR-2003;
Helix Research Institute (JP) ; Research Association for Biotechnology (JP)
FEATURES
source
1. 2684
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 27.5%; Score 396; DB 6; Length 2684;
Best Local Similarity 97.6%; Pred. No. 3.5e-73;
Matches 402; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 1007 TCTGGGTCACTGCTGCAACCGGCGTCTGTCCTTCTTCTGAGAGGCGCTGTGAGTC 1066
DB 2363 TTTCGATGCCCAACCGGCGTCTGTCCTTCTTCTGAGAGGCGCTGTGAGTC 2304
QY 1067 TCCAGTATGTTGGGCGGCGCTTCCGACCCCTTCCGACCAAGCGCCAGAGCTGCA 1126
DB 2303 TCCAGTATGTTGGGCGGCGCTTCCGACCCCTTCCGACCAAGCGCCAGAGCTGCA 2244
QY 1127 GCCAGAGAGAGGGGCGCTCACTCTTATCTTGGGCGACCCACTGCAACAAGGCGCTC 1186
DB 2243 GCCAGAGAGAGGGGCGCTCACTCTTATCTTGGGCGACCCACTGCAACAAGGCGCTC 2184
QY 1187 TCCCAAGCTTAAATGTATCAACCACTACCTGAGAGGGGAGCCCAATCTGACTCTTCC 1246
DB 2183 TCCCAAGCTTAAATGTATCAACCACTACCTGAGAGGGGAGCCCAATCTGACTCTTCC 2124
QY 1247 CCGCCTTGGGACATCGCAGGCGCGGAGAGAGTCCCGGCGGCGCTGAGGCGAGAGCTC 1306
DB 2123 CCGCCTTGGGACATCGCAGGCGCGGAGAGAGTCCCGGCGGCGCTGAGGCGAGAGCTC 2064
QY 1307 CAGGAAGGACACTGAGCGCTGCTGGGCGAGGCGTCCGCAATCCGCAAGGCGCAGAGGAA 1366
DB 2063 CAGGAAGGACACTGAGCGCTGCTGGGCGAGGCGTCCGCAATCCGCAAGGCGCAGAGGAA 2004
QY 1367 GTCTCTGGGGCGATCTGTAATTAACCTTTTCTTTTGTATTTTAAAA 1418
DB 2003 GTCTCTGGGGCGATCTGTAATTAACCTTTTCTTTTGTATTTTAAAA 1952
RESULT 11
AK056896/c 2684 bp mRNA linear PRI 01-AUG-2002
LOCUS AK056896
DEFINITION Homo sapiens cDNA FLJ32334 fis, clone PROSTR2005426.
ACCESSION AK056896
VERSION AK056896.1 GI:16552419
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS
1 Ishibashi, T., Kanehori, K., Yosida, M., Watanabe, S., Ishida, S., Ono, Y., Horita, T., Hiraoka, S., Murakawa, K., Takiguchi, S., Kusano, J., Watanabe, M., Fujimori, K., Tanai, H., Ishida, M., Yamashita, H., Chiba, Y., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Makamatsu, A., Ishii, S., Yamamoto, J. I., Isono, Y., Kawai-Hito, Y., Saito, K., Nishikawa, T., Kimura, K., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Takahashi-Fujii, A.,

Oshtima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuko, Y., Nagai, K. and Isogai, T.
TITLE Unpublished
JOURNAL NEDO human cDNA sequencing project
REFERENCE
AUTHORS
2 (bases 1 to 2684)
Isogai, T., Otsuki, T. and Sugiyama, T.
TITLE Direct Submission
JOURNAL Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
COMMENT
NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology (RAB); cDNA library construction; Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5' & 3' end one pass sequencing; RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing; RAB and HRI.
FEATURES
source
1. 2684
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PROSTR2005426"
/tissue_type="prostate"
/clone_lib="PROST2"
/note="cloning vector: pME18SL3"
56. .952
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB71304.1"
/db_xref="gi:16552420"
/translation="MARTGHPFVYAPKPTFPMWDTLASIMFILATLFTVILPG /RGRRLWLRVYTSLEFLGRIIGTVQQLININNEFTRLGVAEPAKLE /KGLPDPVLYAEKTPSPGCLYRQYLAGYTSAMLVAFLLCMLANVLSPELVY /GYSMLATGIFQLALILFFSATELSGPCPLHGLASVLAITHGPAFWITLTGILCVL /GLMAVAHBMQPHRLKAFPMQSVDEDPMLSEWSEEGGLSPRYRWADSPKSDIPL /SEASTKAYCKEAPHKDPDCAL"
ORIGIN
Query Match 27.5%; Score 396; DB 9; Length 2684;
Best Local Similarity 97.6%; Pred. No. 3.5e-73;
Matches 402; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 1007 TCTGGGTCACTGCTGCAACCGGCGTCTGTCCTTCTTCTGAGAGGCGCTGTGAGTC 1066
DB 2363 TTTCGATGCCCAACCGGCGTCTGTCCTTCTTCTGAGAGGCGCTGTGAGTC 2304
QY 1067 TCCAGTATGTTGGGCGGCGCTTCCGACCCCTTCCGACCAAGCGCCAGAGCTGCA 1126
DB 2303 TCCAGTATGTTGGGCGGCGCTTCCGACCCCTTCCGACCAAGCGCCAGAGCTGCA 2244
QY 1127 GCCAGAGAGAGGGGCGCTCACTCTTATCTTGGGCGACCCACTGCAACAAGGCGCTC 1186
DB 2243 GCCAGAGAGAGGGGCGCTCACTCTTATCTTGGGCGACCCACTGCAACAAGGCGCTC 2184
QY 1187 TCCCAAGCTTAAATGTATCAACCACTACCTGAGAGGGGAGCCCAATCTGACTCTTCC 1246
DB 2183 TCCCAAGCTTAAATGTATCAACCACTACCTGAGAGGGGAGCCCAATCTGACTCTTCC 2124
QY 1247 CCGCCTTGGGACATCGCAGGCGCGGAGAGAGTCCCGGCGGCGCTGAGGCGAGAGCTC 1306
DB 2123 CCGCCTTGGGACATCGCAGGCGCGGAGAGAGTCCCGGCGGCGCTGAGGCGAGAGCTC 2064
QY 1307 CAGGAAGGACACTGAGCGCTGCTGGGCGAGGCGTCCGCAATCCGCAAGGCGCAGAGGAA 1366
DB 2063 CAGGAAGGACACTGAGCGCTGCTGGGCGAGGCGCTCCGCAATCCGCAAGGCGCAGAGGAA 2004
QY 1367 GTCTCTGGGGCGATCTGTAATTAACCTTTTCTTTTGTATTTTAAAA 1418
DB 2003 GTCTCTGGGGCGATCTGTAATTAACCTTTTCTTTTGTATTTTAAAA 1952

RESULT 12
AC009700 156534 bp DNA linear HTG 20-APR-2000
LOCUS AC009700
DEFINITION Homo sapiens chromosome 15 clone RP11-163P10 map 15, WORKING DRAFT
SEQUENCE 15 unordered pieces.
ACCESSION AC009700
VERSION AC009700.4 GI:7622346
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 156534)
Britten, B., Linton, L., Nusbaum, C., and Lander, E.
Homo sapiens chromosome 15, clone RP11-163P10
Unpublished
2 (bases 1 to 156534)
Britten, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baker, J., Baldwin, J., Barna, N., Beckert, R., Benn, J., Brown, A.,
Caestele, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., Dearlano, K., Depyre, E., Devon, K., Dewar, K.,
Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C.,
Funks, R., Gage, D., Galagan, J., Gaidyna, S., Gilbert, D., Grant, G.,
Hagos, B., Heaford, A., Horton, L., Howland, C., Jones, C., Kam, L.,
Karatas, A., Lehotzky, J., Lien, C., Locke, K., Macdonald, P.,
Margulis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J.,
Meidrin, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,
Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,
Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,
Testa, S., Torrella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,
Wheeler, J., Wu, X., Wyman, D., Ye, W., and Zody, M.
Direct Submission
Submitted (28-AUG-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 20, 2000 this sequence version replaced g1:6056272.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: 163_P10
Center clone name: 163_P10
Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 108837 bases at least Q40
Consensus quality: 130230 bases at least Q30
Consensus quality: 145227 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 155134; sum-of-contigs
Quality coverage: 3.2 in Q20 bases; agarose-fp
Quality coverage: 3.3 in Q20 bases; sum-of-contigs
NOTE: This is a 'working draft' sequence. It currently
consists of 15 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1 1169: contig of 1169 bp in length
* 1170 1269: gap of 100 bp
* 1270 3335: contig of 2066 bp in length
* 3336 3435: gap of 100 bp
* 3436 5698: contig of 2263 bp in length
* 5699 5798: gap of 100 bp

FEATURES
source
* 5799 9216: contig of 3418 bp in length
* 9217 9316: gap of 100 bp
* 9317 15633: contig of 6317 bp in length
* 15634 15733: gap of 100 bp
* 15734 21746: contig of 6013 bp in length
* 21747 21846: gap of 100 bp
* 21847 30351: contig of 8505 bp in length
* 30352 30451: gap of 100 bp
* 30452 36509: contig of 6058 bp in length
* 36510 36609: gap of 100 bp
* 36610 45279: contig of 8670 bp in length
* 45280 45379: gap of 100 bp
* 45380 56851: contig of 11472 bp in length
* 56852 56951: gap of 100 bp
* 56952 67374: contig of 10423 bp in length
* 67375 67474: gap of 100 bp
* 67475 79534: contig of 12060 bp in length
* 79535 79634: gap of 100 bp
* 79635 95432: contig of 15798 bp in length
* 95433 95532: gap of 100 bp
* 95533 120801: contig of 25269 bp in length
* 120802 120901: gap of 100 bp
* 120902 156534: contig of 35633 bp in length.
Location/Qualifiers
1. 156534
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-163P10"
/clone_id="RP11-163P10"
1. 1169
/note="assembly_fragment"
1270. 3335
/note="assembly_fragment"
3436. 5698
/note="assembly_fragment"
5799. 9216
/note="assembly_fragment"
9317. 15633
/note="assembly_fragment"
15734. 21746
/note="assembly_fragment"
21847. 30351
/note="assembly_fragment"
30452. 36509
/note="assembly_fragment"
36609. 45279
vector_side:left
36610. 45279
/note="assembly_fragment"
45380. 56851
/note="assembly_fragment"
56952. 67374
/note="assembly_fragment"
67475. 79534
/note="assembly_fragment"
79635. 95432
/note="assembly_fragment"
95533. 120801
/note="assembly_fragment"
120902. 156534
/note="assembly_fragment"

ORIGIN
Query Match 27.5% Score 396; DB 2; Length 156534;
Best Local Similarity 97.6%; Pred. No. 4.6e-73;
Matches 402; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 1007 TCTGGATCAGCTGGCAACCGCGCTGTGCTCTTCTCTGAGAGGCGCTGTGATGTC 1066
Db 96111 TTTCATCCCAACCCCAAGGCGCTGTGCTCTTCTCTGAGAGGCGCTGTGATGTC 96170

QY 1067 TCCAGTATGTTGGCCCGCCGCTCTTGACACCCCTTGACCAAAAGCCGCAAGACTGCA 1126
 Db 96171 TCCAGTATGTTGGCCCGCCGCTCTTGACACCCCTTGACCAAAAGCCGCAAGACTGCA 96230
 QY 1127 GCCAGGAGAGAGGGGGCTCACTCTTAATCTCGGCGACCCCACTGCAAGAGGGGGCTC 1186
 Db 96231 GCCAGGAGAGGGGGCTCACTCTTAATCTCGGCGACCCCACTGCAAGAGGGGGCTC 96230
 QY 1187 TCCAGACTTAAATGATATCACTAACCTGTGAGGGGAGCCCAATCTGACTCTTC 1246
 Db 96291 TCCAGACTTAAATGATATCACTAACCTGTGAGGGGAGCCCAATCTGACTCTTC 96350
 QY 1247 CCGCTTGGGACATCCGAGGCGGGAGAGAGTCCCGCCAGAGCTTGGGCGAGAGAGCTC 1306
 Db 96351 CCGCTTGGGACATCCGAGGCGGGAGAGAGTCCCGCCAGAGCTTGGGCGAGAGAGCTC 96410
 QY 1307 CAGGAGGCGCACTGAGGCTGCTGCGCGAGGCTCGACATCCGAGGCGAGAGAGAA 1366
 Db 96411 CAGGAGGCGCACTGAGGCTGCTGCGCGAGGCTCGACATCCGAGGCGAGAGAGAA 96470
 QY 1367 GTCTCTGGGGCGCATCTGTAAATAACCTTTTCTTTGTTTAAAA 1418
 Db 96471 GTCTCTGGGGCGCATCTGTAAATAACCTTTTCTTTGTTTAAAA 96522

RESULT 13

AX405919 1636 bp DNA linear PAT 14-JUN-2002
 DEFINITION Sequence 334 from Patent WO0222660.
 ACCESSION AX405919
 VERSION AX405919.1 GI:21439277
 KEYWORDS

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1
 TANG, Y.T., LIU, C., ZHOU, P., ASUNDI, V., ZHANG, J., ZHAO, Q.A., REN, F.,
 XUE, A.J., YANG, Y., WEHMAN, T. and DRMANAC, R.T.
 Novel nucleic acids and polypeptides
 Patent: WO 0222660-A 334 21-MAR-2002;
 JOURNAL HYSEQ, INC. (US)

FEATURES
 Location/Qualifiers
 1..1636

CDS
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 441..1472
 /note="unamed protein product"
 /codon_start=1
 /protein_id="CAD35026.1"
 /db_xref="GI:21439278"
 /db_xref="EMBL:CAD35026"
 /translation="MATLGFPEFVAGKPTFPMDTTLLASIIIMIFLALATPVLIPG
 IRGKTRFLMLRLVSTLFGAAILAVNSESEVGVSTNTYKAFSSPMISADIGLO
 VTGSGNNITLTPVQINFTINNEFTMRLEENAEVAAALEGEPDPTVLTKEK
 FTFRSPGGLRQRLAGHTYSMLNVAFTDMLANNMPLVLTGTYLLTGLTGLFOL
 LALLFMSMATSLSVPCPLHGSVLTHTHGPATWITLTLGLCVLLGLAMAVAHMOP
 HRLKAFNOSVDEDDPMLWSPBEGGLSPRYRSMADSPKSDIPISEASTYAYCKEA
 HKKDPDCAL"

ORIGIN

Query Match 24.7% Score 356.4; DB 6; Length 1636;
 Best local similarity 63.9%; Pred. No. 8e-65;
 Matches 556; Conservative 0; Mismatches 311; Indels 3; Gaps 1;

QY 255 CAGCATGACCTCTGGAGAACGGCGTACGCTTTTACCCCAAGCCCGGATGCCGACAG 314
 Db 437 CAAGATGGCTACTTTGGGACACACATCCCTCTTATGTGGCCCAAGCAACCTTCC 496
 QY 315 CTGAGCGTTGCACTGCTCATCGTATTATTCAGTCTTTTGGCTCTAGACAGCAAGCTTCT 374

Db 497 GATGACACCACTTTGGCCAGCATCATCATATCTTTCTGACTGCACCTGGCCAGCTTCAT 556
 QY 375 GCTCATCTTGGCGGGATCCGTGGCCACTCGCGCTGTGTTGGTTGGAGATTCTTCT 434
 Db 557 CGTCATCTGCTGGCATTCGGGAAAGCGAGGCTGTCTGGGCTCTCGGGTGGTAC 616
 QY 435 CAGCTGTTCAATAGGCGAGAAATGTGGCTGTGACCTTCACTGAGAGATGTTGTTGG 494
 Db 617 CAGCTTATTCATCGGGGCTGCAATCTGAGCTGTGAATTTCACTTGTGAGTGTCTGTGG 676
 QY 495 TACAGTAAACCAACACATCTTCAAAAGCCCTTACGCGAGCGCGGTAAAGCCGTGT 554
 Db 677 CAGGTCAAGACCAACATCATTAAGAGCCCTTCACTGAGTGTGATGACGCTGATAT 736
 QY 555 CGGCTGCTGTGGGCTGTGAGGCGATTAATTTACATCAAGAGAGCCCAAGTATCA 614
 Db 737 TGGGCTGAGGTGGGCTGGGTGGATCAATCACTCAAGAGAGCCCGGTGACGA 796
 QY 615 GCTGAACGAGACCATGACTATCAAGAGAGCTTCACTGCGCTTGAAGAGAAATTAAGC 674
 Db 797 GCTGAATGAGACCATGATTAACAAGAGAGTTCACTGGGCGCTGGGTGAATATAGC 856
 QY 675 CGCGAGGTACGGAAGCACTGAGAGAGGCGCTGCCGAGCCAGTGTCTTACCTGGCGGA 734
 Db 857 TGAGGAGTATGCAAAAGGCTGTGAGAGAGGCGTGCAGACCTGTGTGTACTTAATGA 916
 QY 735 GAAGTTCACACGAGTACGAGCCCTTGGCGCTGTATACACAGTACAGTACCTGGCGGACACTA 794
 Db 917 GAAGTTCATCCCAAGAGGCCATGTGGCTTAATACCGCAGATACCGCTGGGGGACACTA 976
 QY 795 CGGCTGGCCAGCTATAGGAGTGGCTTGTGCTTGTGCTCTCTTCAAGCTGTCTCTC 854
 Db 977 CACCTCAGCCATCTATAGGAGTGGCATCTCTGTGCTGGCTGGCTGGCAATGTGATCTCTC 1036
 QY 855 CAGCGGCGCCCGCTGTACGAGAGCGCTGACATGCTGACACACGAGAGCCCTTGGCGCTCT 914
 Db 1037 CAGCTGCGCTGTATATGT 1096
 QY 915 CG---GGGTTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 971
 Db 1097 GGCTGTGCTTCTTCTTCATGAGCCACATCACTCACTCACTCACTCACTCACTCACTCACT 1156
 QY 972 CTGCTCGGCGCTACCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 1031
 Db 1157 CGCTTGT 1216
 QY 1032 CTGTGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1091
 Db 1217 GCTGT 1276
 QY 1092 TCGCACCTTCTTGGACCAAGCGCCAGGA 1121
 Db 1277 GAAAGCTTCTTCAACCAAGGTGTGATGA 1306

RESULT 14

BC029819 1923 bp mRNA linear PRI 06-OCT-2003
 DEFINITION Homo sapiens hypothetical protein FLJ22334, mRNA (cDNA clone
 MGC:55310 IMAGE:5177619), complete cds.
 ACCESSION BC029819
 VERSION BC029819.1 GI:20987583
 KEYWORDS MGC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1923)
 STRAUSBERG, R.D., FEINGOLD, E.A., GROUSE, L.H., DERGE, J.G.,
 KLAUSNER, R.D., COLLINS, F.S., WAGNER, L., SHERMAN, C.M., SCHULTE, G.D.,
 ALTSCHUL, S.F., ZEEBERG, B., BUETOW, K.H., SCHAEFER, C.F., BHAT, N.K.,
 HOPKINS, R.F., JORDAN, H., MOORE, T., MAX, S.I., WANG, J., HATTEH, F.,
 DATCHENKO, L., MATSUNA, K., FARMER, A.A., RUDIN, G.M., HONG, L.,

Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
 Scheetz, T.E., Brownstein, M.J., Udell, T.B., Toshiyuki, S.,
 Scarinci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
 Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J.,
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
 Fahy, U., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
 Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shechenko, Y.,
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smalins, D.E.,
 Scherch, A., Schein, J.E., Jones, S.J. and Marr, M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL MEDLINE 22388257
 12477932
 2 (bases 1 to 1923)

REFERENCE PubMed
 12477932

AUTHORS Strausberg, R.
 TITLE Direct Submission

JOURNAL Submitted (06-MAY-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabbs@mail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louisse, H.,
 Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nantavali,
 A.N., Gibbs, R.A.

FEATURES
 source
 1..1923
 Location/Qualifiers
 1..1923
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:35310 IMAGE:517619"
 /tissue_type="Brain, lung, testis, adult, pooled whole"
 /clone_id="NIH_MGC_115"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 1..1923
 /gene="FLJ32334"
 /db_xref="LOCUSID:90527"
 401..1852
 /codon_start=1
 /product="Flj32334 protein"
 /protein_id="AAH29819.1"
 /db_xref="GI:20987584"
 /db_xref="LOCUSID:90527"
 /translation="MATLIGTPFVAGPKPTPMDTLLASIMIFLATATFVILPG
 IRGTRFLMILRYVTSIFGAILAVPSSSEMSVQSVNTSYKAFSSPWSADIGLQ
 VEGSVNITITGTPVOQNETINNEBFTRRIGENAEYAKALEKGLDPDLYIAEK
 FTPEPSCGLRQYRLAGHYTSAMLVAFDMLVAFVMSIPVYGGYMLATGIFOL
 LALFFMSQVSLSPCPHLHAGVLIHTHGPAPWITLTLGLCVLDGAAVAHROP
 HRLKAFNQSVDPEPMLSESPBEGGLSPRYRMSADSPKSDIPLSEASTAVYRPR
 RLISLPAVIRGLAFAALASALPGALLQAWMALPGHRCRKAQESRLGPHSPMRFEP
 EGCEERVAERTGDSFRLRGRTGLRMWMSKERRACGVRAMLPRLVNSGIXKRPSCL
 DLPKCMYRRDARAFHLEPPTCVTSRHPTPL"

gene
 CDS
 401..1852

ORIGIN		Query Match	24.7%	Score 356.4	DB 9	Length 1923	
Beet local similarity		63.9%	Pred. No. 8.1e-65				
Matches 556		Consecutive	0	Mismatches 311	Indels 3	Gaps 1	
QY	255	CAGCATACCCCTGTGAAACGGCGTACTGCTTTTACCCCGAGCCCGGATGCGCAG	314				
DB	337	CAAGATGCTACTTTGGGACACACATCCCTCTATGCTGGGCCCAAGCCATCTCC	456				
QY	315	CTTACGGCTTCACTGCTCATCTGTTATCTAGTCTTTTGGCTTACAGCAAGCTTCT	374				
DB	457	GATGACACCACTTTGGCCAGCATCATGATCTTTCTGACTGCACTGGCCAGCTTCA	516				
QY	375	GCTCATCTTGGCCGGGATCCGTGGCCATCGCCGCTGTTTGGTTGGTAGAGTTCT	434				
DB	517	CGTCATCTGCTGCTGGCATTTGGGGAAAGAGAGGCTGTTCTGCTGCTTGGGTTG	576				
QY	435	CAGTCTGTTCAATAGCGCGAATATTTGCTGTGCACTTACAGTACAGAAATGCTGG	494				
DB	577	CAGCTTATTCATCGGGGCTGCATCTGCTGTGAATTCAGTTCTGAGTGTCTGG	636				
QY	495	TACGTGAACACCAACATCTCTCAAGGCTTACAGCGCAGCGGTTACAGCCGTG	554				
DB	637	CCAGTTCAGACCAACACATATACAGGCTTCAAGTTCAGATGATCAGCGCTATAT	696				
QY	555	CGGTCTGCTGTGGGCTGTGAGGCAATTAATTTACCTACAGAGAACCCAGTGCAT	614				
DB	697	TGGCTGAGGCTGGGCTGGGCTGGTGGATGATCACTACAGAGGACCCCGTGGCA	756				
QY	615	GCTGACAGACCACTGATCTCAAGAGAGTTCACCTGGCGTCTGAAAGATTAAGC	674				
DB	757	GCTGATGATGACATCAATTAACAGAGAGTTCACCTGGCGTCTGAAAGATTAAGC	816				
QY	675	CGCGAGTACCGGAAACGACTGGAAGAGGCTGCGGACCACTGCTTACTGCGGGA	734				
DB	817	TGAGAGTATGAAAGGCTGTGAGAGGCTGCGGACCACTGCTTACTGCGGGA	876				
QY	735	GAGTTCACACCGAGTACCTTGGGCTGTACACAGATACCACTGGCGGAGACTA	794				
DB	877	GAGTTCACCAAGACCAATGAGCTTATACCGCAAGTACCGCTGGCGGAGACTA	936				
QY	795	CGCTTCGACGAGTATGAGTGTGCTTGTGCTCTTCCAAAGTGTGCTCTC	854				
DB	937	CACCTACGACATGATATGAGTGTGCTCTTGTGCTCTTCCAAAGTGTGCTCTC	996				
QY	855	CAGCGGCGCCGCTCTAGAGAGGCTGGCACTGCTACACAGGAGCTTGGGCTCT	914				
DB	997	CAGCTGTGCTGTATATGAGGCTATATGAGGCACTGCTACAGGAGCTTGGGCTCT	1056				
QY	915	CG--GAGTCTTGGCTTGGCTTCCATCTAGCGTGGCGCTTGGCGGCTCGGCTAG	971				
DB	1057	GAGTGTGCTTCTTCTCATGAGGCACTACCTCACCCGTGCTCGGCACTGGG	1116				
QY	972	CTCTTCGCGCTACCACTCAGTACGGGCGGCTTGGGCTACCGCTGCAACGCGGT	1031				
DB	1117	CGTCTGTGCTGCACTACCACTCAGTACGGGCGGCTTGGGCTACCGCTGCAACG	1176				
QY	1032	CTGTGCTCTTCTTGGAGGCGGCTGAGTCTCCAGTATGTTGGCCAGCGCTCT	1091				
DB	1177	GCTGTGTGCTGCTGAGGCTGCTGATAGGCGGTGCGCACAGATGAGCTTCAAGCT	1236				
QY	1092	TGCGACCCCTTGGACCAAGGCCAAGA	1121				
DB	1237	GAAAGCTTTCTTCAACCAAGTGTGATGA	1266				

RESULT 15
 BC019755
 LOCUS BC019755 1521 bp mRNA linear ROD 04-OCT-2003
 DEFINITION Mus musculus cDNA sequence BC019755, mRNA (cDNA clone MGC:25550
 IMAGE:3964798), complete cds.
 ACCESSION BC019755

VERSION BC019755.1 GI:18043594
 KEYWORDS MGC.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 REFERENCE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 AUTHORS 1 (bases 1 to 1521)
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klauener, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smallus, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

REMARK
 COMMENT
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulsged, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Navavati, A.N., Gibbs, R.A.

FEATURES
 source
 1..1521
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="MGC:25550 IMAGE:3964798"
 /tissue_type="Mammary tumor, Metastatic adenocarcinoma, TGF alpha model, 10 month old virgin mouse. Taken by biopsy."
 /clone_id="NCI CGAP_Mam1"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 1..1521
 /gene="BC019755"
 /note="synonym: MGC25550"
 /db_xref="locusID:213696"

ORIGIN
 Query Match 24.2%; Score 348.8; DB 10; Length 1521;
 Best Local Similarity 63.5%; Pred. No. 3.2e-63;
 Matches 549; Conservative 0; Mismatches 312; Indels 3; Gaps 1;
 CDS
 /db_xref="NCI:2384861"
 285..1310
 /codon_start=1
 /product="cDNA sequence BC019755"
 /protein_id="AAH19755.1"
 /db_xref="GI:18043595"
 /db_xref="locusID:213696"
 /translation="MAAGHTLPYTGCTKPEPMDTAVITITLALVTFILPGL
 LRGKRLFWLRVTSFTGAVILAVNPFSSWSVGHANNTTYABSPKVASDVG
 IGLGVNVTFTGTPVQQLNETITNNEAFAMGLVLEAEKALKEKLPVLYLAK
 FLPBSPCLVNYQRLAGHYASAMVAFGLVLANVLSPELVYGGHMLATGLP
 LALPFSMTSLISPCPLRTGLTAVITTHGPAFWITLATGLLCTLLGVAVARMP
 HKLAFNLSSDPVLWESSEGLSPHYSIASEPDTQTPMSVASSETCFEEDP
 KESDSL"

QY 279 ACTGCTTTTACCCGAGCCCGGACGAGGCTTACGCTTCACCTGCTCATGCT 338
 DB 305 ACTGCTTTTACCCGAGCCCGGACGAGGCTTACGCTTCACCTGCTCATGCT 364
 QY 339 TATTCAGTGTGTTTGGCTTACGAGCAAGCTTCTGCTCATGCTTCCGGGAACTGCTG 398
 DB 365 TATTCAGTGTGTTTGGCTTACGAGCAAGCTTCTGCTCATGCTTCCGGGAACTGCTG 424
 QY 399 CCACCTGCTGTTTGGCTTACGAGCAAGCTTCTGCTCATGCTTCCGGGAACTGCTG 458
 DB 425 GAAAGAGAGCTGTTTGGCTTACGAGCAAGCTTCTGCTCATGCTTCCGGGAACTGCTG 484
 QY 459 TGTGCTGTGACCTTACGAGCAAGCTTCTGCTCATGCTTCCGGGAACTGCTG 518
 DB 485 CTGCTGTGAGCAAGCTTACGAGCAAGCTTCTGCTCATGCTTCCGGGAACTGCTG 544
 QY 519 CAAAGCTTACGAGCAAGCTTCTGCTCATGCTTCCGGGAACTGCTG 578
 DB 545 CAAAGCTTACGAGCAAGCTTCTGCTCATGCTTCCGGGAACTGCTG 604
 QY 579 CATTATATTTACCTACGAGCAAGCTTCTGCTCATGCTTCCGGGAACTGCTG 638
 DB 605 CATTATATTTACCTACGAGCAAGCTTCTGCTCATGCTTCCGGGAACTGCTG 664
 QY 639 CGAGCAAGCTTACCTGCTTACGAGCAAGCTTCTGCTCATGCTTCCGGGAACTGCTG 698
 DB 665 TGAGGCAATGCAATGAGGCTTACGAGCAAGCTTCTGCTCATGCTTCCGGGAACTGCTG 724
 QY 699 GAAAGGAGCTGCGGACCCGAGGCTTACGAGCAAGCTTCTGCTCATGCTTCCGGGAACTGCTG 758
 DB 725 GAAAGGAGCTGCGGACCCGAGGCTTACGAGCAAGCTTCTGCTCATGCTTCCGGGAACTGCTG 784
 QY 759 CGAGCTTACCTGCTTACGAGCAAGCTTCTGCTCATGCTTCCGGGAACTGCTG 818
 DB 785 TGAGCTTACCTGCTTACGAGCAAGCTTCTGCTCATGCTTCCGGGAACTGCTG 844
 QY 819 GTTGTGCTTACCTGCTTACGAGCAAGCTTCTGCTCATGCTTCCGGGAACTGCTG 878
 DB 845 ATTCTGCTTACCTGCTTACGAGCAAGCTTCTGCTCATGCTTCCGGGAACTGCTG 904
 QY 879 CTTGAGCTTACCTGCTTACGAGCAAGCTTCTGCTCATGCTTCCGGGAACTGCTG 935
 DB 905 CCAATGCTTACCTGCTTACGAGCAAGCTTCTGCTCATGCTTCCGGGAACTGCTG 964
 QY 936 CATCTGAGCTTACCTGCTTACGAGCAAGCTTCTGCTCATGCTTCCGGGAACTGCTG 995
 DB 965 CACCTGAGCTTACCTGCTTACGAGCAAGCTTCTGCTCATGCTTCCGGGAACTGCTG 1024
 QY 996 CGAGGAGCTTACCTGCTTACGAGCAAGCTTCTGCTCATGCTTCCGGGAACTGCTG 1055
 DB 1025 TGAGGAGCTTACCTGCTTACGAGCAAGCTTCTGCTCATGCTTCCGGGAACTGCTG 1084
 QY 1056 CGTGTGAGCTTACCTGCTTACGAGCAAGCTTCTGCTCATGCTTCCGGGAACTGCTG 1115

Db 1085 CATGGCAGTGGCCCAAGAGATGACGCTCAGAGCTGAAGGCTTTCTTCACCTGAGTTC 1144
QY 1116 CAAGGACTGCAGCCAGAGAGAGG 1139
Db 1145 AGAGGACCCAGTGTGAGTGGG 1168

Search completed: February 23, 2004, 21:57:50
Job time : 5691.97 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 19:33:52 ; Search time 582.643 Seconds

(without alignments) 10506.704 Million cell updates/sec

Title: US-09-936-456-1

Perfect score: 1441

Sequence: 1 aaagtaacgctcagacagc.....aaaaaaaaaaaaaaaa 1441

Scoring table: IDENTITY_NUC

Gapop 10.0 ; Gapext 1.0

Searched: 337863 seqs, 2124099041 residues 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq_29Jan04:*

1: geneseqn1980s:.*
2: geneseqn1990s:.*
3: geneseqn2000s:.*
4: geneseqn2001as:.*
5: geneseqn2001bs:.*
6: geneseqn2002s:.*
7: geneseqn2003as:.*
8: geneseqn2003bs:.*
9: geneseqn2003cs:.*
10: geneseqn2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1441	100.0	1441	AAA94623	AA94623 Human CAS
2	1436	99.7	1491	ADD19228	ADD19228 Human CDN
3	1435.6	99.6	1474	AB190535	AB190535 Human pol
4	1413	98.1	1421	AAA95605	AA95605 CDNA enco
5	1392.8	96.7	1460	AA95564	AA95564 Human sec
6	1392.8	96.7	1460	AB267235	AB267235 Human sec
7	1392.8	96.7	1460	AB273640	AB273640 Secreted
8	1392.8	96.7	1460	ADC20289	ADC20289 Human sec
9	1392.8	96.7	1460	AAA94624	AA94624 Human CAS
10	1392.8	96.7	1460	AB268115	AB268115 Human sec
11	1392.8	96.7	1460	AB274587	AB274587 Secreted
12	1392.8	96.7	1460	ADC21005	ADC21005 Human sec
13	1392.8	96.7	1460	ADD19187	ADD19187 Human CDN
14	1392.8	96.7	1460	AAA95627	AA95627 Human cod
15	1392.8	96.7	1460	AB267235	AB267235 Human sec
16	1392.8	96.7	1460	AB273640	AB273640 Secreted
17	1392.8	96.7	1460	ADC20289	ADC20289 Human sec
18	1392.8	96.7	1460	AAA94624	AA94624 Human CAS
19	1392.8	96.7	1460	AB268115	AB268115 Human sec
20	1392.8	96.7	1460	AB274587	AB274587 Secreted
21	1392.8	96.7	1460	ADC21005	ADC21005 Human sec
22	1392.8	96.7	1460	ADD19187	ADD19187 Human CDN
23	1392.8	96.7	1460	AAA95627	AA95627 Human cod

24	65.6	4.6	1602	4	AB121213	AB121213 Drosophila
25	65.6	4.6	1808	4	AB107775	AB107775 Drosophila
26	64	4.4	512	4	AAH36185	AAH36185 Human col
27	60.6	4.2	2000	7	ADA71938	ADA71938 Rice gene
28	52.6	3.7	134	7	AB268116	AB268116 Human sec
29	52.6	3.7	134	7	AB274588	AB274588 Secreted
30	52.6	3.7	134	7	ADC21006	ADC21006 Human sec
31	52	3.6	2000	7	ADA71938	ADA71938 Rice gene
32	49	3.4	496	5	ABV56664	ABV56664 Human pro
33	48.6	3.4	1400	4	AA801017	AA801017 Maize dis
34	47.8	3.3	396	3	AA234971	AA234971 Wheat bet
35	47.6	3.3	277	7	ABX47508	ABX47508 Bovine ES
36	47.4	3.3	915	3	AA60068	AA60068 Human sec
37	46.8	3.2	456	4	AA123653	AA123653 Human bre
38	46.8	3.2	2484	5	AA578178	AA578178 DNA enco
39	46.8	3.2	2589	7	AB275977	AB275977 Human G p
40	46.8	3.2	14055	7	AA161170	AA161170 Actinosyr
41	46.8	3.2	82746	7	AA161224	AA161224 Actinosyr
42	46.4	3.2	2206	4	ABK43595	ABK43595 DNA enco
43	46.2	3.2	441	3	AA696600	AA696600 Human sec
44	46.2	3.2	531	8	ADA49299	ADA49299 Maize gen
45	46.2	3.2	609	8	ADA49313	ADA49313 Maize gen

ALIGNMENTS

RESULT 1

ID: AAA94623 standard; DNA; 1441 BP.

AAA94623;

11-JAN-2001 (first entry)

Human CASB618 coding sequence.

Epitope; human; CASB618; chromosome 15; tumour; vaccine; cancer; ovarian;

colon; autoimmune disease; HLA_A0201; ss.

CS Homo sapiens.

XX Key Location/Qualifiers

FT CDS /tag= a /product= "Human CASB618"

PD WC0200053748-A2.

14-SEP-2000.

XX 09-MAR-2000; 2000WC-EP002048.

XX 11-MAR-1999; 99GB-00005607.

XX 01-SEP-1999; 99GB-00020590.

XX (SMIK) SWITKLINE BEECHAM BIOLOGICALS.

XX Bruck CEM, Casart J, Coche T, Vinals Y De BassolsC;

XX WPI; 2000-572268/53.

XX P-PSDB; AAB26325.

XX New human CASB618 polypeptide, useful as a vaccine for prophylactic and

XX therapeutic treatment of cancer, particularly ovarian or colon cancer,

XX autoimmune diseases and related conditions.

XX Claim 13; Page 61; 76pp; English.

XX The present sequence is the coding sequence of human CASB618 protein. The

XX gene for human CASB618 is thought to be located on chromosome 15. The

XX protein encoded by the present sequence and epitopes of the CASB618

XX protein (see AAB26327 to AAB26399) are useful in diagnosing the occurrence

CC of tumour cells and in vaccines for prophylactic and therapeutic
 CC treatment of cancers, particularly ovarian or colon cancer, autoimmune
 CC diseases and related conditions
 XX

SQ Sequence 1441 BP; 289 A; 466 C; 392 G; 294 T; 0 U; 0 Other;

Query Match 100.0%; Score 1441; DB 3; Length 1441;
 Best Local Similarity 100.0%; Pred. No. 1,3e-311;
 Matches 1441; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 AAAGTACGGCTTACAGACAGTGAAGAAATGTTGGCTTGGCCGCTTGAAGAAATCTGTG 60
DB 1 AAAGTACGGCTTACAGACAGTGAAGAAATGTTGGCTTGGCCGCTTGAAGAAATCTGTG 60
QY 61 GTACCAACCCCGAGAGGTTGAGAGAGCCCACTTCCAGCTTCTTAAAGAGAGTGA 120
DB 61 GTACCAACCCCGAGAGGTTGAGAGAGCCCACTTCCAGCTTCTTAAAGAGAGTGA 120
QY 121 GGAATCAGACTTACACAGCCCACTGGTCCAGCTTGTACGCAAGAGAGCCAGAC 180
DB 121 GGAATCAGACTTACACAGCCCACTGGTCCAGCTTGTACGCAAGAGAGCCAGAC 180
QY 181 GGGCTCTCCGGCTCCAGAGAGCCCACTGGTCTGAGTCCGCTCCGCTGGTGGAGC 240
DB 181 GGGCTCTCCGGCTCCAGAGAGCCCACTGGTCTGAGTCCGCTCCGCTGGTGGAGC 240
QY 241 ACTGCGCCGCGCTGAGCATGACCTGTGGAACGCGCTACTGCTTTTACCCCAAGCC 300
DB 241 ACTGCGCCGCGCTGAGCATGACCTGTGGAACGCGCTACTGCTTTTACCCCAAGCC 300
QY 301 CGGATGCGCGAGGCTTCAAGCTTCCACTGCTCACTGCTTATTTAGTGTGGCTCTA 360
DB 301 CGGATGCGCGAGGCTTCAAGCTTCCACTGCTCACTGCTTATTTAGTGTGGCTCTA 360
QY 361 GCAGCAAGCTTCTGCTCACTTGTGCGGGGATCCGTGGCCACTGCGCTGGTTTGGTTG 420
DB 361 GCAGCAAGCTTCTGCTCACTTGTGCGGGGATCCGTGGCCACTGCGCTGGTTTGGTTG 420
QY 421 GTGAGAGTCTTCTCACTGCTTGTTCATAGGCGGAGAAATTTGGCTGTGCACTTACGTGA 480
DB 421 GTGAGAGTCTTCTCACTGCTTGTTCATAGGCGGAGAAATTTGGCTGTGCACTTACGTGA 480
QY 481 GAATGCTTGTGGGTACAGTGAACCAACCAACATCTTCAAAAGCTTCAAGCGAGCCGCG 540
DB 481 GAATGCTTGTGGGTACAGTGAACCAACCAACATCTTCAAAAGCTTCAAGCGAGCCGCG 540
QY 541 GTTACAGCCCGGTGTGCTGCTGTGGGCTTGAAGGGATTAATATTAACAACAAGGG 600
DB 541 GTTACAGCCCGGTGTGCTGCTGTGGGCTTGAAGGGATTAATATTAACAACAAGGG 600
QY 601 ACCCGAGTGCATCAGTGAACGAGACCATTAAGTCAACAGAGCACTTCACTGGAGTCTG 660
DB 601 ACCCGAGTGCATCAGTGAACGAGACCATTAAGTCAACAGAGCACTTCACTGGAGTCTG 660
QY 661 AAAGAGATTACCGCGCGAGTACGGAACGCACTGAGAGAGGGCTGCCGAGCCCAAGTG 720
DB 661 AAAGAGATTACCGCGCGAGTACGGAACGCACTGAGAGAGGGCTGCCGAGCCCAAGTG 720
QY 721 CTGTACCTGGCGAGAGATTCAACCGAGTAGCCCTTGGCGCTTACCAACAGTACAC 780
DB 721 CTGTACCTGGCGAGAGATTCAACCGAGTAGCCCTTGGCGCTTACCAACAGTACAC 780
QY 781 CTGGCGGAGACATACGCTTGGCGAGCTATAGGGTGGCTTGTGCTTGTGCTCTCTCC 840
DB 781 CTGGCGGAGACATACGCTTGGCGAGCTATAGGGTGGCTTGTGCTTGTGCTCTCTCC 840
QY 841 AAGTGTGCTCTTCCAGCGCGCGCTCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
DB 841 AAGTGTGCTCTTCCAGCGCGCGCTCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
QY 901 GCGTTGCGCTCTTGGGGGTCTTGGCTTGGCTTGCATCTAGAGGTGCGGCTGCGCG 960
DB 901 GCGTTGCGCTCTTGGGGGTCTTGGCTTGGCTTGCATCTAGAGGTGCGGCTGCGCG 960

```

```

QY 961 CTCGCGCTAAGGCTCTCCGCGCTACACCACTAGTACGGGCGCGCTTGGGTACGCTG 1020
DB 961 CTCGCGCTAAGGCTCTCCGCGCTACACCACTAGTACGGGCGCGCTTGGGTACGCTG 1020
QY 1021 GCAACCGGCGCTCTGTGCTCTTCTTCCGAGAGGGGCGGTGGTGTCTCCAGTATGTTCCG 1080
DB 1021 GCAACCGGCGCTCTGTGCTCTTCTTCCGAGAGGGGCGGTGGTGTCTCCAGTATGTTCCG 1080
QY 1081 CCGAGCGCTCTTGGCAACCTTCTGAGACCAAGGAGGCAAGGAGTGCAGGAGAGAGAGG 1140
DB 1081 CCGAGCGCTCTTGGCAACCTTCTGAGACCAAGGAGGCAAGGAGTGCAGGAGAGAGAGG 1140
QY 1141 GAGTCACTCTTATCTCTGCGAGACCACTGCAACAGAGCGCTCTCCAGACTTAA 1200
DB 1141 GAGTCACTCTTATCTCTGCGAGACCACTGCAACAGAGCGCTCTCCAGACTTAA 1200
QY 1201 TGTATCACTATACCTGTGAGGGGAGCCCACTGTGACTCTTCCCGCTTGGGACAT 1260
DB 1201 TGTATCACTATACCTGTGAGGGGAGCCCACTGTGACTCTTCCCGCTTGGGACAT 1260
QY 1261 CGCAGGCGGAGACAGTGCCTGCGCAGGCTGAGGCGCAGAGACTCCAGAGAGGACATG 1320
DB 1261 CGCAGGCGGAGACAGTGCCTGCGCAGGCTGAGGCGCAGAGACTCCAGAGAGGACATG 1320
QY 1321 AGCGCTGCTGGCGGAGGCTCTGAGACATCCGACAGGACCCAGGAAAGTCTCTGGGGCA 1380
DB 1321 AGCGCTGCTGGCGGAGGCTCTGAGACATCCGACAGGACCCAGGAAAGTCTCTGGGGCA 1380
QY 1381 TCTGTAATTAACCTTTTCTTTTCTTTTAAAAAATTTTAAAAAATTTTAAAAA 1440
DB 1381 TCTGTAATTAACCTTTTCTTTTCTTTTAAAAAATTTTAAAAAATTTTAAAAA 1440
QY 1441 A 1441
DB 1441 A 1441

RESULT 2
ADD19228
ID ADD19228 standard; cDNA; 1491 BP.
XX
AC ADD19228;
XX
DT 15-JAN-2004 (first entry)
XX
DE Human cDNA from secreted protein gene 45.
XX
KW human secreted protein; cytosolic; antibacterial; virucide;
KW neuroprotective; gynaecological; gastrointestinal; Gen; cardiant;
KW cardiovascular; Gen; nephrologic; antiinflammatory; muscular; Gen;
KW respiratory; Gen; immunosuppressive; cerebroprotective; vasotropic;
KW neotropic; antiallergic; cancer; bacterial infection; viral infection;
KW neural disorder; immune system disorder; blood disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW inflammatory disorder; proliferative disorder; human; ss; gene.
XX
OS Homo sapiens.
XX
PN W02003052377-A2.
XX
PD 26-JUN-2003.
XX
PF 06-NOV-2002; 2002WO-US035606.
XX
PR 07-NOV-2001; 2001US-0331046P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-533050/50.

```

DR P-PSDB; ABD19303.

XX New isolated nucleic acids encoding signal transduction pathway component
PT polypeptides, useful for diagnosing, treating, and/or preventing
PT disorders, such as cancer, infections, cardiovascular and inflammatory
PT diseases.

XX Claim 1; SEQ ID NO 55; 554bp; English.

XX The invention relates to an isolated nucleic acid molecule (cDNA)
CC encoding a human secreted protein, representing one of 85 novel genes.
CC Also included are recombinant vectors, host cells (expressing the
CC protein), the secreted proteins (including their fragments, epitopes and
CC homologues), an isolated antibody that binds specifically to the protein,
CC diagnosing a pathological condition or susceptibility to a pathological
CC condition (comprising determining the presence or absence of a mutation
CC in the nucleic acid and diagnosing a condition based on the presence or
CC absence of the mutation), diagnosing a pathological condition or
CC susceptibility to a pathological condition (comprising determining the
CC presence or amount of expression of the protein in a biological sample
CC and diagnosing a condition based on the presence or amount of expression
CC of the protein), preventing, treating or ameliorating a medical condition
CC by administering the nucleic acid or protein to a mammalian subject,
CC identifying a binding partner to the protein, the gene corresponding to
CC the cDNA sequence, and identifying an activity in a biological assay
CC (comprising expressing the nucleic acid in a cell, isolating the
CC supernatant, detecting an activity in a biological assay and identifying
CC the protein in the supernatant having the activity). The nucleic acids
CC and proteins display the following activities: Cytostatic, antibacterial,
CC Vmucide, Neuroprotective, Gynaecological, Gastrointestinal, Gen,
CC Cardiant, Cardiovascular-Gen, Nephrotropic, Antiinflammatory, Muscular-
CC Neurotropic, Antiatherogenic. The methods and compositions of the present
CC invention are useful for diagnosing, treating, preventing and/or
CC prognosticating disorders related to the novel polypeptides, such as
CC cancer, bacterial or viral infections, and neural, immune system, blood,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal, inflammatory or proliferative disorders (many examples of these
CC diseases and disorders are given in the specification). The present
CC sequence encodes a novel secreted protein of the invention.

XX Sequence 1491 BP; 293 A; 485 C; 410 G; 302 T; 0 U; 1 Other;

Query Match 99.7%; Score 1436; DB 9; Length 1491;

Best Local Similarity 99.9%; Pred. No. 1.7e-310; Mismatches 1; Indels 0; Gaps 0;

Matches 1436; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAGTAAAGGCTTCAAGACAGTGAATAGTTTGGCTGCGGCTAAGAAAATCTGTGCG 60
DB 44 AAGTAAAGGCTTCAAGACAGTGAATAGTTTGGCTGCGGCTAAGAAAATCTGTGCG 103
QY 61 GTACCAACCCAGAGGCTTGAAGACAGCCACCTTCCAGCTTAAAGAGAGTGA 120
DB 104 GTACCAACCCAGAGGCTTGAAGACAGCCACCTTCCAGCTTAAAGAGAGTGA 163
QY 121 GAGTACAGATTACACAGCCCACTGGTCCAGCTTGAAGCAAGAGAGCCAAAGAC 180
DB 164 GAGTACAGATTACACAGCCCACTGGTCCAGCTTGAAGCAAGAGAGCCAAAGAC 223
QY 181 GCGCTCTCCGCGCTTCAAGAGAGCCCAAGCTTGTGCTTGTGCTGCGGCTGCGAGC 240
DB 224 GCGCTCTCCGCGCTTCAAGAGAGCCCAAGCTTGTGCTTGTGCTGCGGCTGCGAGC 283
QY 241 ACTGCGCGGCGTGAAGCATGACCTGTGGAAGCGGTAAGCTTTTAAACCCCAAGCC 300
DB 284 ACTGCGCGGCGTGAAGCATGACCTGTGGAAGCGGTAAGCTTTTAAACCCCAAGCC 343
QY 301 CGGCAATCCGAGAGCTTCAAGCTTCAAGCTTCAATGATTTTGTGCTTGA 360
DB 344 CGGCAATCCGAGAGCTTCAAGCTTCAAGCTTCAATGATTTTGTGCTTGA 403
QY 361 CGAGCAAGCTTCTGCTCATCTTGGCGGAGATCCGTGCGCACTCGCGCTGTTGTG 420

DB 404 CGAGCAAGCTTCTGCTCATCTTGGCGGAGATCCGTGCGCACTCGCGCTGTTGTG 463
QY 421 GTGAGAGTCTTCTCAGTCTGTGTAAGAGCGCAAGAAATGGCTGTGCACTTCAAGTGA 480
DB 464 GTGAGAGTCTTCTCAGTCTGTGTAAGAGCGCAAGAAATGGCTGTGCACTTCAAGTGA 523
QY 481 GAATGATTTGTGGGTGAAGTGAACCAACCAATCTTAACAAAGCTTCAAGCGAGCGCGC 540
DB 524 GAATGATTTGTGGGTGAAGTGAACCAACCAATCTTAACAAAGCTTCAAGCGAGCGCGC 583
QY 541 GTTACAGCCCGTGTGCTGTGCTGTGGGCGCTGAAGGGCAATTAATTAATCACTCAAGG 600
DB 584 GTTACAGCCCGTGTGCTGTGCTGTGGGCGCTGAAGGGCAATTAATTAATCACTCAAGG 643
QY 601 ACCCAAGTCACTAAGCTGAAGCAAGCAATTAAGTCAACAGAGAGTTCACTGAGCGTGTG 660
DB 644 ACCCAAGTCACTAAGCTGAAGCAAGCAATTAAGTCAACAGAGAGTTCACTGAGCGTGTG 703
QY 661 AAGAGAAATTACCGCGCGAGTGAAGCGCACTGAAGAGGGGCTGCGGAGCCCAAGT 720
DB 704 AAGAGAAATTACCGCGCGAGTGAAGCGCACTGAAGAGGGGCTGCGGAGCCCAAGT 763
QY 721 CTCTACCTGCGGAGAGTTCACACCGAGTACCTTGGCGCTGTACCACTAGTACAC 780
DB 764 CTCTACCTGCGGAGAGTTCACACCGAGTACCTTGGCGCTGTACCACTAGTACAC 823
QY 781 CTGCGGAGACATACGCTTGGCGCAAGCTATAGGTGAGCTTGTGCTTGTGCTTCTTCC 840
DB 824 CTGCGGAGACATACGCTTGGCGCAAGCTATAGGTGAGCTTGTGCTTGTGCTTCTTCC 883
QY 841 AACGTGCTCTTCCACCGCGCGCGCTTCAAGAGAGCGCTGAGCACTGATCAACACGGA 900
DB 884 AACGTGCTCTTCCACCGCGCGCGCTTCAAGAGAGCGCTGAGCACTGATCAACACGGA 943
QY 901 GCTTGTGCGCTTGTGCGGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCGG 960
DB 944 GCTTGTGCGCTTGTGCGGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCGG 1003
QY 961 CTGCGGAGTGTGCTTGTGCGGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCGG 1020
DB 1004 CTGCGGAGTGTGCTTGTGCGGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCGG 1063
QY 1021 GCAACCGGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCGG 1080
DB 1064 GCAACCGGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCGG 1123
QY 1081 CCGAGGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCGG 1140
DB 1124 CCGAGGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCGG 1183
QY 1141 GGTCACTCTTATCTGTGCGGCAAGCTTGAAGCAAGAGCGGCTTCCAGACTTAA 1200
DB 1184 GGTCACTCTTATCTGTGCGGCAAGCTTGAAGCAAGAGCGGCTTCCAGACTTAA 1243
QY 1201 TGTATCAACACTTAACCTGTGAAGGGGAGCCCAATCTGAGCTTCCCGCTTGGGACAT 1260
DB 1244 TGTATCAACACTTAACCTGTGAAGGGGAGCCCAATCTGAGCTTCCCGCTTGGGACAT 1303
QY 1261 CGAGGCGCGGAGAGTGCAGCGGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCGG 1320
DB 1304 CGAGGCGCGGAGAGTGCAGCGGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCGG 1363
QY 1321 AGCGCTGTGCGGAGAGCTTGAAGCATTCGCAAGAGCACTAGAGAAAGTCTCTTGGGCGA 1380
DB 1364 AGCGCTGTGCGGAGAGCTTGAAGCATTCGCAAGAGCACTAGAGAAAGTCTCTTGGGCGA 1423
QY 1381 TCTGTAAATAAAGCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1438
DB 1424 TCTGTAAATAAAGCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1481

RESULT 3
ABL90535

ID ABL90535 standard; cDNA; 1474 BP.
XX ABL90535;
AC
XX 24-MAY-2002 (first entry)
DT
XX Human polynucleotide SEQ ID NO 1097.
DE
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; antitumor;
XX vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein; gene; ss.
OS Homo sapiens.
XX
XX MO200190304-A2.
XX
XX 29-NOV-2001.
XX
XX 18-MAY-2001; 2001WO-US016450.
XX
XX 19-MAY-2000; 2000US-020551SP.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Birse CE, Rosen CA;
PI
XX WPI; 2002-122018/16.
DR P-PSDB; ABB90126.
XX
XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT prevention of neural, immune system, muscular, reproductive,
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT disorders.
XX
XX Claim 4; SEQ ID NO 1097; 2081bp + Sequence Listing; English.
PS
XX
XX The invention relates to novel genes (ABL99449-ABL90853) and proteins
CC (ABB99040-ABB90444) useful for preventing, treating or ameliorating
CC isolated conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing
CC / (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pat_sequences
XX
XX
SQ Sequence 1474 BP; 292 A; 480 C; 404 G; 296 T; 0 U; 2 Other;

Query Match 99.6%; Score 1435.6; DB 6; Length 1474;
Best Local Similarity 99.8%; Pred. No. 2,1e-310;
Matches 145; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAGTAAAGGCTTACAGACAGTGAAGAAATAGTTTGGCTGCGCGGCTAGAAAAAAGCTTGTG 60
DB 27 AAAGTAAAGGCTTACAGACAGTGAAGAAATAGTTTGGCTGCGCGGCTAGAAAAAAGCTTGTG 86
QY 61 GTACCAACCCAGAGGCTTACAGACAGTGAAGAAATAGTTTGGCTGCGCGGCTAGAAAAAAGCTTGTG 120
DB 87 GTACCAACCCAGAGGCTTACAGACAGTGAAGAAATAGTTTGGCTGCGCGGCTAGAAAAAAGCTTGTG 146
QY 121 GGAATCAAGCTTACAGACAGTGAAGAAATAGTTTGGCTGCGCGGCTAGAAAAAAGCTTGTG 180
DB 147 GGAATCAAGCTTACAGACAGTGAAGAAATAGTTTGGCTGCGCGGCTAGAAAAAAGCTTGTG 206

QY 181 GGGCTTCGCGCGGCTTACAGACAGTGAAGAAATAGTTTGGCTGCGCGGCTAGAAAAAAGCTTGTG 240
DB 207 GGGCTTCGCGCGGCTTACAGACAGTGAAGAAATAGTTTGGCTGCGCGGCTAGAAAAAAGCTTGTG 266
QY 241 ACTGCGCGCGGCTTACAGACAGTGAAGAAATAGTTTGGCTGCGCGGCTAGAAAAAAGCTTGTG 300
DB 267 ACTGCGCGCGGCTTACAGACAGTGAAGAAATAGTTTGGCTGCGCGGCTAGAAAAAAGCTTGTG 326
QY 301 CGGATGCGCGGCTTACAGACAGTGAAGAAATAGTTTGGCTGCGCGGCTAGAAAAAAGCTTGTG 360
DB 327 CGGATGCGCGGCTTACAGACAGTGAAGAAATAGTTTGGCTGCGCGGCTAGAAAAAAGCTTGTG 386
QY 361 GAGCAAGCTTCTGATCATCTTCCGCGGAGATCGGCGGCTAGAAAAAAGCTTGTG 420
DB 387 GAGCAAGCTTCTGATCATCTTCCGCGGAGATCGGCGGCTAGAAAAAAGCTTGTG 446
QY 421 GTGAGAGTTCTTCTGATCATCTTCCGCGGAGATCGGCGGCTAGAAAAAAGCTTGTG 480
DB 447 GTGAGAGTTCTTCTGATCATCTTCCGCGGAGATCGGCGGCTAGAAAAAAGCTTGTG 506
QY 481 GAAATGCTTCCGCGGCTTACAGACAGTGAAGAAATAGTTTGGCTGCGCGGCTAGAAAAAAGCTTGTG 540
DB 507 GAAATGCTTCCGCGGCTTACAGACAGTGAAGAAATAGTTTGGCTGCGCGGCTAGAAAAAAGCTTGTG 566
QY 541 GTTACAGCGGCTTCTGATCATCTTCCGCGGAGATCGGCGGCTAGAAAAAAGCTTGTG 600
DB 567 GTTACAGCGGCTTCTGATCATCTTCCGCGGAGATCGGCGGCTAGAAAAAAGCTTGTG 626
QY 601 ACCGAGAGTATGAGTGAAGAAATAGTTTGGCTGCGCGGCTAGAAAAAAGCTTGTG 660
DB 627 ACCGAGAGTATGAGTGAAGAAATAGTTTGGCTGCGCGGCTAGAAAAAAGCTTGTG 686
QY 661 AAAGAAATTAAGCGCGGAGTACAGACAGTGAAGAAATAGTTTGGCTGCGCGGCTAGAAAAAAGCTTGTG 720
DB 687 AAAGAAATTAAGCGCGGAGTACAGACAGTGAAGAAATAGTTTGGCTGCGCGGCTAGAAAAAAGCTTGTG 746
QY 721 CTCTACCTGCGGAGAAATTAAGCGCGGAGTACAGACAGTGAAGAAATAGTTTGGCTGCGCGGCTAGAAAAAAGCTTGTG 780
DB 747 CTCTACCTGCGGAGAAATTAAGCGCGGAGTACAGACAGTGAAGAAATAGTTTGGCTGCGCGGCTAGAAAAAAGCTTGTG 806
QY 781 CTGCGGAGTATGAGTGAAGAAATAGTTTGGCTGCGCGGAGTACAGACAGTGAAGAAATAGTTTGGCTGCGCGGCTAGAAAAAAGCTTGTG 840
DB 807 CTGCGGAGTATGAGTGAAGAAATAGTTTGGCTGCGCGGAGTACAGACAGTGAAGAAATAGTTTGGCTGCGCGGCTAGAAAAAAGCTTGTG 866
QY 841 AAGTGTGCTTCTGATCATCTTCCGCGGAGATCGGCGGCTAGAAAAAAGCTTGTG 900
DB 867 AAGTGTGCTTCTGATCATCTTCCGCGGAGATCGGCGGCTAGAAAAAAGCTTGTG 926
QY 901 GCGTTCGCGGCTTCTGATCATCTTCCGCGGAGATCGGCGGCTAGAAAAAAGCTTGTG 960
DB 927 GCGTTCGCGGCTTCTGATCATCTTCCGCGGAGATCGGCGGCTAGAAAAAAGCTTGTG 986
QY 961 CTGCGGAGTATGAGTGAAGAAATAGTTTGGCTGCGCGGAGTACAGACAGTGAAGAAATAGTTTGGCTGCGCGGCTAGAAAAAAGCTTGTG 1020
DB 987 CTGCGGAGTATGAGTGAAGAAATAGTTTGGCTGCGCGGAGTACAGACAGTGAAGAAATAGTTTGGCTGCGCGGCTAGAAAAAAGCTTGTG 1046
QY 1021 GCAACCGGCGGCTTCTGATCATCTTCCGCGGAGATCGGCGGCTAGAAAAAAGCTTGTG 1080
DB 1047 GCAACCGGCGGCTTCTGATCATCTTCCGCGGAGATCGGCGGCTAGAAAAAAGCTTGTG 1106
QY 1081 CCGAGCGGCTTCTGATCATCTTCCGCGGAGATCGGCGGCTAGAAAAAAGCTTGTG 1140
DB 1107 CCGAGCGGCTTCTGATCATCTTCCGCGGAGATCGGCGGCTAGAAAAAAGCTTGTG 1166
QY 1141 GCGTCAAGCTTCTGATCATCTTCCGCGGAGATCGGCGGCTAGAAAAAAGCTTGTG 1200
DB 1167 GCGTCAAGCTTCTGATCATCTTCCGCGGAGATCGGCGGCTAGAAAAAAGCTTGTG 1226
QY 1201 TGTATCAAGCTTCTGATCATCTTCCGCGGAGATCGGCGGCTAGAAAAAAGCTTGTG 1260
DB 1227 TGTATCAAGCTTCTGATCATCTTCCGCGGAGATCGGCGGCTAGAAAAAAGCTTGTG 1286

QY 1261 CGAGAGCCGGGAAGACAGTCCCGCCAGAGCCTGGGCGCAGAGAGCTCCAGGAAGGCGACTG 1320
DB 1287 CGCAGGCCGGGAAGACAGTCCCGCCAGAGCCTGGGCGCAGAGAGCTCCAGGAAGGCGACTG 1346
QY 1321 AGGCGTCTGCGCGGAGGCTCGGACATCCGCGAGGACCAAGGAAAGTCTCTGGGGCGGA 1380
DB 1347 AGCGCTGCTGCGCGGAGGCTCGGACATCCGCGAGGACCAAGGAAAGTCTCTGGGGCGGA 1406
QY 1381 TCTGTAATAAACCCTTTTCTTTGTTTAAATAAAAAAAAAAAAAAAAAAAAAA 1438
DB 1407 TCTGTAATAAACCCTTTTCTTTGTTTAAATAAAAAAAAAAAAAAAAAAAAAA 1464

RESULT 4
AAA96505
ID AAA96505 standard; cDNA; 1421 BP.

XX AAA96505;

DT 08-FEB-2001 (first entry)

DE cDNA encoding a human transmembrane protein.

KM Human; transmembrane protein; cell proliferation disorder; myeloma;
KW reproductive disorder; smooth muscle disorder; neurological disorder;
KW arteriosclerosis; leukaemia; acquired immunodeficiency syndrome; AIDS;
KW allergy; ovulatory defect; angina; hypertension; stroke; epilepsy;
KW Alzheimer's disease; Tourette's disorder; ss.

OS Homo sapiens.

EH Key Location/Qualifiers
FT CDS 267..1229
FT /tag= a

PN W0200056891-A2.

XX 28-SEP-2000.

PF 22-MAR-2000; 2000NC-US007817.

XX 22-MAR-1999; 99US-0125537P.

PR 16-JUN-1999; 99US-0139565P.

XX (INCY-) INCYTE PHARM INC.

PI Yue H, Lal P, Tang YT, Hillman JL, Reddy R, Bandman O;
PI Baughn MR, Lu DM, Azimzai Y, Yang J;

DR WPI; 2000-579485/54.
P-PSDB; AAB18992.

PT New human transmembrane proteins are used to treat a disease or condition
PT associated with decreased expression of functional HTMP e.g. Tourette's
PT disorder, angina and leukemia.

XX Claim 4; Page 129; 130pp; English.

PS The present sequence encodes a human transmembrane proteins (HTMP).
CC Agonists and antagonists of the protein are used to treat a disease or
CC condition associated with overexpression of the protein. Diseases and
CC conditions which can be treated include cell proliferative,
CC immunological, reproductive, smooth muscle and neurological disorders
CC e.g. arteriosclerosis, myeloma, leukaemia, acquired immunodeficiency
CC syndrome (AIDS), allergies, ovulatory defects, angina, hypertension,
CC stroke, Alzheimer's disease, epilepsy and Tourette's disorder. The
CC polynucleotides may be used to detect and quantify gene expression in
CC biopsied tissues where protein expression may be correlated with disease
CC e.g. to determine absence/ presence or excess expression of HTMP or to
CC monitor regulation of HTMP expression during therapeutic intervention
SQ Sequence 1421 BP; 264 A; 468 C; 395 G; 294 T; 0 U; 0 Other;

Query Match 98.1%; Score 1413; DB 3; Length 1421;
Best Local Similarity 100.0%; Pred. No. 2,3e-305;
Matches 1413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAGTAAAGGGCTTACAGACAGTGAAGAAATAGTTTGGCTTCGGCGCTAGAAAAAATCTGTG 60
DB 9 AAAGTAAAGGGCTTACAGACAGTGAAGAAATAGTTTGGCTTCGGCGCTAGAAAAAATCTGTG 68
QY 61 GTACCAACCCCGAGAGGTTGAGAGAGGCCACCTGCACAGCTTCCCTTAAACGAGAGGTGCA 120
DB 69 GTACCAACCCCGAGAGGTTGAGAGAGGCCACCTGCACAGCTTCCCTTAAACGAGAGGTGCA 128
QY 121 GGAAGTCAAGCTTACCAAGCCCACTGGTCCAGACCTTGTACGCAAAAGAGCCAGAGAC 180
DB 129 GGAAGTCAAGCTTACCAAGCCCACTGGTCCAGACCTTGTACGCAAAAGAGCCAGAGAC 188
QY 181 GGGCTCTCCGGGTCAGAGGACGCCACCTGTGGCTGGCTGGCGGCGGAGTGAAGC 240
DB 189 GCGCTTCCCGGTCAGAGGACGCCACCTGTGGCTGGCTGGCGGCGGAGTGAAGC 248
QY 241 ACTGCGCGCGGTGAGCATGACCTCTGTGAACGCGGTACTGCTTTTACCCCGAGCC 300
DB 249 ACTGCGCGCGGTGAGCATGACCTCTGTGAACGCGGTACTGCTTTTACCCCGAGCC 308
QY 301 CGGCAATGCCGAGGCTTACAGCTTCCACTGCTCATCGTTATTTAGTATTTTGGCTCTA 360
DB 309 CGGCAATGCCGAGGCTTACAGCTTCCACTGCTCATCGTTATTTAGTATTTTGGCTCTA 368
QY 361 GCAGCAAGCTTCTGCTCATCTTGGCGGGGATCCGTGCGCACTCGCGTGGTTTGGTTG 420
DB 369 GCAGCAAGCTTCTGCTCATCTTGGCGGGGATCCGTGCGCACTCGCGTGGTTTGGTTG 428
QY 421 GTGAGAGTCTTCTCATGCTGTTCAADAGCGGAAATTTGGCTGTGCACTTCAAGTGA 480
DB 429 GTGAGAGTCTTCTCATGCTGTTCAADAGCGGAAATTTGGCTGTGCACTTCAAGTGA 488
QY 481 GAATGTTGCTGGGTGATAGTGAACCAACACATCTTCAAAAGCCTTACGCGAGCGCGC 540
DB 489 GAATGTTGCTGGGTGATAGTGAACCAACACATCTTCAAAAGCCTTACGCGAGCGCGC 548
QY 541 GTTACAGCCCGTGTGCTGTGCTGTGGGCGCTGAGGGGCAATTAATTAATTAATTAATTA 600
DB 549 GTTACAGCCCGTGTGCTGTGCTGTGGGCGCTGAGGGGCAATTAATTAATTAATTAATTA 608
QY 601 ACCCGAGTCAATGAGTGAAGAGACCAATTGATCAACAGAGAGTCACTGAGCGGTG 660
DB 609 ACCCGAGTCAATGAGTGAAGAGACCAATTGATCAACAGAGAGTCACTGAGCGGTG 668
QY 661 AAAGAGAAATTAAGCGCGCGAGTACCGGAAAGCACTGAGAAAGGGCTTCCGAGCCAGTG 720
DB 669 AAAGAGAAATTAAGCGCGCGAGTACCGGAAAGCACTGAGAAAGGGCTTCCGAGCCAGTG 728
QY 721 CTTACTTGGCGGAGAAATTCACACGAGTGAAGCTTGGGCTGTATACACAGTATACAC 780
DB 729 CTTACTTGGCGGAGAAATTCACACGAGTGAAGCTTGGGCTGTATACACAGTATACAC 788
QY 781 CTGGCGGAGCACTACAGCTCGGCGACGAGTATGAGGTGAGTGTGCTGTGCTCTCC 840
DB 789 CTGGCGGAGCACTACAGCTCGGCGACGAGTATGAGGTGAGTGTGCTGTGCTCTCC 848
QY 841 AAGTGTCTCTCTTCAAGCGCGGCGCGCTTACGAGAGGCTTGCATCTCAACACAGGGA 900
DB 849 AAGTGTCTCTCTTCAAGCGCGGCGCGCTTACGAGAGGCTTGCATCTCAACACAGGGA 908
QY 901 GCGTTGCGGCTCTTGGGAGTCTTGGGCTTGTGAGCTTCACTTGAAGTGCCTGTGCGG 960
DB 909 GCGTTGCGGCTCTTGGGAGTCTTGGGCTTGTGAGCTTCACTTGAAGTGCCTGTGCGG 968
QY 961 CTCGCGCTTGGGCTCTTGGGCTTGTGAGCTTCACTTGAAGTGCCTGTGCGGCTG 1020
DB 969 CTCGCGCTTGGGCTCTTGGGCTTGTGAGCTTCACTTGAAGTGCCTGTGCGGCTG 1028
QY 1021 GCAACCGGCGTCTGTGCTTCTCTGGAAGGGGCGGTGAGTGTCAAGTATATTTGCG 1080

```

Db      1029 GCACCGGCGTCTGCTGCTTCTCCGAGGCGCCGAGTGTCCAGTATGTTCCG 1088
QY      1081 CCCAGGCTCTTGAGACCTTGTGACCAAGGCGCAAGCTGACGAGGAGGAGG 1140
Db      1089 CCCAGGCTCTTGAGACCTTGTGACCAAGGCGCAAGCTGACGAGGAGGAGG 1148
QY      1141 GGCTACCTCTTATCTGCGGCGACCACTGACCAAGGCGCTCTCCAGATTAAA 1200
Db      1149 GGCTACCTCTTATCTGCGGCGACCACTGACCAAGGCGCTCTCCAGATTAAA 1208
QY      1201 TGTATCACACTTAACCTGTGAGGAGGAGGACCACTGACCTCTTCCCGCTTGGGACAT 1260
Db      1209 TGTATCACACTTAACCTGTGAGGAGGAGGACCACTGACCTCTTCCCGCTTGGGACAT 1268
QY      1261 CGCAGGCGCGGAGAGAGTGCCTCCGACGAGCTGTGAGGAGGAGTCCAGAGGAGGAGCTG 1320
Db      1269 CGCAGGCGCGGAGAGAGTGCCTCCGACGAGCTGTGAGGAGGAGTCCAGAGGAGGAGCTG 1328
QY      1321 AGCGTGTGCGGCGGAGGAGCTTGACATCCGACGAGGAGGAGGAGGAGTCTCTGGGCGCA 1380
Db      1329 AGCGTGTGCGGCGGAGGAGCTTGACATCCGACGAGGAGGAGGAGGAGTCTCTGGGCGCA 1388
QY      1381 TCTGTAAATAAACCTTTTCTTTTCTTTT 1413
Db      1389 TCTGTAAATAAACCTTTTCTTTTCTTTT 1421

```

RESULT 5
AAC95564
ID AAC95564 standard; cDNA; 1460 BP.

XX AAC95564;

DT 21-FEB-2001 (first entry)

DE Human secreted protein gene 44 SEQ ID NO:54.

XX Human; secreted protein; cytosolic; immunosuppressive; nocrotropic;
 XX neuroprotective; antiviral; antiallergic; hepatotropic; antibacterial;
 XX antiinflammatory; antitumor; anticonvulsant; antibacterial;
 XX antifungal; antiparasitic; cardiac; cancer; immune disease; allergy;
 XX cardiovascular disorder; wound healing; infection; neurological disease;
 XX ss.

XX Homo sapiens.

XX MO200061596-A1.

XX 19-OCT-2000.

XX 06-APR-2000; 2000MO-US008983.

XX 09-APR-1999; 99US-0128703P.

XX 14-JAN-2000; 2000US-0176068P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (ROSE/) ROSEN C A.

XX Ruben SM, Komatsoulis G;

XX WPI: 2000-611865/56.

XX P-PSDB; AAB52055.

XX Fifty nucleic acid molecules encoding human secreted proteins, useful in

XX the prevention, treatment and diagnosis of cancer, immune disorders,

XX cardiovascular disorders and neurological diseases.

XX Claim 1; Page 443; 505pp; English.

XX Polynucleotide sequences AAC95521 - AAC95570 represent cDNA encoding 50

XX human secreted proteins AAB52012 - AAB52161. Sequences AAB52062 -

XX AAB52103 represent alternative polypeptides encoded by the genes, and

CC amino acid sequences with which they share homology. The genes and
 CC proteins have activities dependent on the tissues and cells in which they
 CC are expressed. Examples of their activities include cytosolic;
 CC immunosuppressive; nocrotropic; neuroprotective; antiviral; antiallergic;
 CC hepatotropic; antidiabetic; antiinflammatory; antitumor; anticonvulsant;
 CC antiparasitic; antiallergic; antiparasitic; and cardiac.
 CC The secreted proteins, polypeptides, antagonists and agonists may be
 CC useful in treating, preventing and/or diagnosing diseases and disorders
 CC such as cancer, particularly breast and ovarian cancer, and other cancers
 CC of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract,
 CC liver, lung, or urogenital. Immune disorders such as Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; cardiovascular disorders such as
 CC myocardial ischaemia; wound healing; neurological diseases such as
 CC cerebral anoxia and epilepsy; and infectious diseases such as viral,
 CC bacterial, fungal and parasitic infections may also be treated using the
 CC proteins and polynucleotides of the invention. Sequences AAC95512 -
 CC AAC95520 and AAB52011 are used in the isolation and characterisation of
 CC the proteins and polynucleotides of the invention
 XX

Sequence 1460 BP; 292 A; 472 C; 398 G; 298 T; 0 U; 0 Other;

Query Match 96.7%; Score 1392.8; DB 3; Length 1460;
 Best Local Similarity 99.3%; Pred. No. 7.5e-301;
 Matches 1430; Conservative 0; Mismatches 7; Indels 3; Gaps 3;

```

QY      1 AAAGTAACGCTACAGACAGTGAAGTAAGTTTGCTGCGCGGCTAGAAAACTCTGTG 60
Db      22 AAAGTAACGCTACAGACAGTGAAGTAAGTTTGCTGCGCGGCTAGAAAACTCTGTG 81
QY      61 GTACCAACCCGAGGCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120
Db      82 GTACCAACCCGAGGCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 141
QY      121 GGACTCAGACTTACACAGCCCACTCGGTCCAGCTTTGTAAGCAAGAGAGGCAAGAGAG 180
Db      142 GGACTCAGACTTACACAGCCCACTCGGTCCAGCTTTGTAAGCAAGAGAGGCAAGAGAG 201
QY      181 GGGCTCTCCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 240
Db      202 GGGCTCTCCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 261
QY      241 ACTCGGCGGCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
Db      262 ACTCGGCGGCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 321
QY      301 CGGATGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360
Db      322 CGGATGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 381
QY      361 GCAGCAAGCTTCTGCTCATCTTTCGCGGAGATCCGTTGCGGAGGAGGAGGAGGAGGAGG 420
Db      382 GCAGCAAGCTTCTGCTCATCTTTCGCGGAGATCCGTTGCGGAGGAGGAGGAGGAGGAGG 441
QY      421 GTGAGAGTTCTTTCAGTCTGTTTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 480
Db      442 GTGAGAGTTCTTTCAGTCTGTTTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 501
QY      481 GAATGTTCTGTTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 540
Db      502 GAATGTTCTGTTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 561
QY      541 GTTACAGCCCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
Db      562 GTTACAGCCCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 621
QY      601 ACCCAGGAGTACGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 660
Db      622 ACCCAGGAGTACGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 681
QY      661 AAGAGAAATTAAGCCGCGAGTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 720

```

Db 682 AAGAGATTACGCGGAGTACGGAACGACTGAGAAAGGGCTGCCGACCCACTG 741
 Qy 721 CTCTACCTGGCGGAGAGTTACACCGAGTACCTTGGCGCTTACACCACTGACAC 780
 Db 742 CTCTACCTGGCGGAGAGTTACACCGAGTACCTTGGCGCTTACACCACTGACAC 801
 Qy 781 CTGGCGGAGACATACGCTGCGGACGACTAGGGTGGCTTGGCTTGGCTTGGCTTGG 840
 Db 802 CTGGCGGAGACATACGCTGCGGACGACTAGGGTGGCTTGGCTTGGCTTGGCTTGG 861
 Qy 841 AACGTGCTCTTCCACGCGCGCCGCTCTACGAGAGCTTGGACCTGCTGACACCGGA 900
 Db 862 AACGTGCTCTTCCACGCGCGCCGCTCTACGAGAGCTTGGACCTGCTGACACCGGA 921
 Qy 901 GCGCTGCGCTCTTGGCGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 960
 Db 922 GCGCTGCGCTCTTGGCGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 981
 Qy 961 CTCCGCTTACGCTCTTCCGCGCTACACCTACGTAAGCGCGCTTCTGGCTGACGCTG 1020
 Db 982 CTCCGCTTACGCTCTTCCGCGCTACACCTACGTAAGCGCGCTTCTGGCTGACGCTG 1041
 Qy 1021 GCAACGCGCTCTTCCGCTCTTCCGCGCTCTTCCGCGCTCTTCCGCGCTCTTCCGCGCT 1080
 Db 1042 GCAACGCGCTCTTCCGCTCTTCCGCGCTCTTCCGCGCTCTTCCGCGCTCTTCCGCGCT 1101
 Qy 1081 CCGAGCGCTCTTCCGCGCTCTTCCGCGCTCTTCCGCGCTCTTCCGCGCTCTTCCGCGCT 1139
 Db 1102 CCGAGCGCTCTTCCGCGCTCTTCCGCGCTCTTCCGCGCTCTTCCGCGCTCTTCCGCGCT 1161
 Qy 1140 GGGGTACCTCTTATCTCTGCGGACACCTGACCAAGCGCGCTCTTCCGCGCTCTTCCGCGCT 1198
 Db 1162 GGGGTACCTCTTATCTCTGCGGACACCTGACCAAGCGCGCTCTTCCGCGCTCTTCCGCGCT 1220
 Qy 1199 AATGTATCACCACTTATCTGAGGGGACCCCATCTGACTCTTCCGCGCTTGGGAC 1258
 Db 1221 AATGTATCACCACTTATCTGAGGGGACCCCATCTGACTCTTCCGCGCTTGGGAC 1280
 Qy 1259 ATGCGAGCGCGGAGAGAGTGGCGCGCGCTTGGGACCGAGAGTCTCCAGAGAGGAC 1318
 Db 1281 ATGCGAGCGCGGAGAGAGTGGCGCGCGCTTGGGACCGAGAGTCTCCAGAGAGGAC 1340
 Qy 1319 TGAGCGCTCTGCGGAGAGCTCGACATCCGACGACCAAGGAAAGTCTCTGGGAC 1378
 Db 1341 TGAGCGCTCTGCGGAGAGCTCGACATCCGACGACCAAGGAAAGTCTCTGGGAC 1400
 Qy 1379 GATCTGTAATTAACCTTTTCTTTTGTGTTTTTAAAAA 1438
 Db 1401 GATCTGTAATTAACCTTTTCTTTTGTGTTTTTAAAAA 1460
 RESULT 6
 ID ABZ67235 standard, cDNA, 1460 BP.
 XX ABZ67235;
 AC ABZ67235;
 XX 26-MAR-2003 (first entry)
 DE Human secreted protein encoding cDNA SEQ ID NO 355.
 XX
 XX Human; secreted protein; neurotrophic; neuroprotective; cytostatic;
 KW vitruclide; dermatological; immunosuppressive; anti-inflammation; anti-HIV;
 KW vunerary; anticancer; antiparkinsonian; antiskinning; antianemic;
 KW anti-inflammation; cancer; antineumatic; hepatotropic; cerebroprotective;
 KW anti-inflammation; anti-inflammation; antidiabetic; antitumor; anticonvulsant;
 KW antifungal; antiparasitic; cardiac; immune disorder; infection; vaccine;
 KW cardiovascular disorder; neurological disease; nephrotropic;
 KW gene therapy; gene; chromosome 9p21; ds.
 XX Homo sapiens.
 XX
 XX
 PV WO200277186-A2.

XX 03-OCT-2002.
 PD 26-MAR-2002; 2002WO-US009188.
 XX 27-MAR-2001; 2001US-0278650P.
 PR 12-SEP-2001; 2001US-00950082.
 PR 12-SEP-2001; 2001US-00950083.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Rosen CA, Ruben SM;
 PI
 PI WPI; 2003-040583/03.
 DR P-PSDB; ABP99814.
 DR
 PT New human secreted proteins encoded by genes contained in cDNA clones
 PT (e.g. HGCAC19), useful for preventing, treating or diagnosing e.g. AIDS,
 PT multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or
 PT West Nile fever.
 PS Claim 7; Page 1353; 2423dp; English.
 CC The invention relates to novel human genes (ABZ66891-ABZ68209) and the
 CC encoded secreted proteins (ABP99470-ABP99814) useful for preventing,
 CC treating or ameliorating medical conditions e.g. by protein or gene
 CC therapy. The genes are isolated from a range of human tissues disclosed
 CC in the specification. The nucleic acids, proteins, antibodies and
 CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
 CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the
 CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
 CC lung or urogenital; (b) immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
 CC myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g.
 CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
 CC bacterial, fungal and parasitic infections
 XX
 SQ Sequence 1460 BP; 292 A; 472 C; 398 G; 298 T; 0 U; 0 Other;
 Query Match 96.7%; Score 1392.8; DB 7; Length 1460;
 Best Local Similarity 99.3%; Pred. No. 7.5e-301;
 Matches 1430; Conservative 0; Mismatches 7; Indels 3; Gaps 3;
 Qy 1 AAGTAAAGGCTTACAGACGTAAGTAATGTTGCTGCGCGCTAAGAAACTCTGCG 60
 Db 22 AAGTAAAGGCTTACAGACGTAAGTAATGTTGCTGCGCGCTAAGAAACTCTGCG 81
 Qy 61 GTACCAACCCAGAGGCTTACAGACGTAAGTAATGTTGCTGCGCGCTAAGAAACTCTGCG 120
 Db 82 GTACCAACCCAGAGGCTTACAGACGTAAGTAATGTTGCTGCGCGCTAAGAAACTCTGCG 141
 Qy 121 GGACTCAAGCTTACAGACGTAAGTAATGTTGCTGCGCGCTAAGAAACTCTGCG 180
 Db 142 GGACTCAAGCTTACAGACGTAAGTAATGTTGCTGCGCGCTAAGAAACTCTGCG 201
 Qy 181 GCGCTCTCCGCGCTTACAGACGTAAGTAATGTTGCTGCGCGCTAAGAAACTCTGCG 240
 Db 202 GCGCTCTCCGCGCTTACAGACGTAAGTAATGTTGCTGCGCGCTAAGAAACTCTGCG 261
 Qy 241 ACTGCGCGCGCTTACAGACGTAAGTAATGTTGCTGCGCGCTAAGAAACTCTGCG 300
 Db 262 ACTGCGCGCGCTTACAGACGTAAGTAATGTTGCTGCGCGCTAAGAAACTCTGCG 321
 Qy 301 CCGCATGCGCGAGGCTTACAGACGTAAGTAATGTTGCTGCGCGCTAAGAAACTCTGCG 360
 Db 322 CCGCATGCGCGAGGCTTACAGACGTAAGTAATGTTGCTGCGCGCTAAGAAACTCTGCG 381
 Qy 361 GCAGCAAGCTTCTGCTATCTTGGCGGAGATCGTGGCACTCGCGCTGTTGGTTG 420
 Db 382 GCAGCAAGCTTCTGCTATCTTGGCGGAGATCGTGGCACTCGCGCTGTTGGTTG 441

QY 421 GTGAGATCTTCTCAGTCTGTTCATAGGCGCAGAAATGTGCTGTGCACTTCAGTGCA 480
 DB 442 GTGAGATCTTCTCAGTCTGTTCATAGGCGCAGAAATGTGCTGTGCACTTCAGTGCA 501
 QY 461 GATGAGTCTGTGAGTACAGTGAACCAACCACTCTTCAAGGCGGAGGCGG 540
 DB 502 GATGAGTCTGTGAGTACAGTGAACCAACCACTCTTCAAGGCGGAGGCGG 561
 QY 541 GTTACAGCCCGGTGTGCTGTGCTGTGAGGCGGAGGCGGAGGCGGAGGCGG 600
 DB 562 GTTACAGCCCGGTGTGCTGTGCTGTGAGGCGGAGGCGGAGGCGGAGGCGG 621
 QY 601 ACCCGAGTGCATGACCTGAAACGAGCACTTGAACCAAGAGGCGGAGGCGGAGGCGG 660
 DB 622 ACCCGAGTGCATGACCTGAAACGAGCACTTGAACCAAGAGGCGGAGGCGGAGGCGG 681
 QY 661 AAAGAGATTAACCGCGGAGGAGTACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 720
 DB 682 AAAGAGATTAACCGCGGAGGAGTACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 741
 QY 721 CTCTACCTGCGGAGAGTTCACACCGAGTACCTTGTGCGCTGTGACACAGTACAC 780
 DB 742 CTCTACCTGCGGAGAGTTCACACCGAGTACCTTGTGCGCTGTGACACAGTACAC 801
 QY 781 CTGCGGCGGACACTAGGCTCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 840
 DB 802 CTGCGGCGGACACTAGGCTCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 861
 QY 841 AACGTGCTGCTCTCCACGCGGCGGCGGCTCTACGAGAGGCGTGGACCTGACACCGGA 900
 DB 862 AACGTGCTGCTCTCCACGCGGCGGCGGCTCTACGAGAGGCGTGGACCTGACACCGGA 921
 QY 901 GCGTTGCGGCTCTTGGGAGTCTTGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 960
 DB 922 GCGTTGCGGCTCTTGGGAGTCTTGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 981
 QY 961 CTCGCGCTAGGCTCTTGGGAGTCTTGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 1020
 DB 982 CTCGCGCTAGGCTCTTGGGAGTCTTGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 1041
 QY 1021 GCAACCGGCGTCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 1080
 DB 1042 GCAACCGGCGTCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 1101
 QY 1081 CCGAGGCGTCTTGGAGCCTTGTGAGCAAAAGCGGCGGAGGAGGAGGAGGAGGAGGAGG 1139
 DB 1102 CCGAGGCGTCTTGGAGCCTTGTGAGCAAAAGCGGCGGAGGAGGAGGAGGAGGAGGAGG 1161
 QY 1140 GGGCTCAGCTTATCTTCTGCGGACCACTGACAGAGGCGGCTCTCCAGAC-TTAA 1198
 DB 1162 GGGCTCAGCTTATCTTCTGCGGACCACTGACAGAGGCGGCTCTCCAGAC-TTAA 1220
 QY 1199 AATGTATCACTAATCACTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1258
 DB 1221 AATGTATCACTAATCACTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1280
 QY 1259 ATGCAAGCGCGGAGAGCACTGCGGCGGAGGCGGAGGCGGAGGAGGAGGAGGAGGAGGAGG 1318
 DB 1281 ATGCAAGCGCGGAGAGCACTGCGGCGGAGGCGGAGGCGGAGGAGGAGGAGGAGGAGGAGG 1340
 QY 1319 TGAAGCGCTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1378
 DB 1341 TGAAGCGCTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1400
 QY 1379 GATCTGTAAATAAACCCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1438
 DB 1401 GATCTGTAAATAAACCCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1460

RESULT 7
 ABZ73640
 ID ABZ73640 standard; cDNA; 1460 BP.
 XX

AC ABZ73640;
 XX 12-MAY-2003 (first entry)
 DT
 XX
 DE Secreted protein-encoding gene 360 cDNA clone HUFCL31, SEQ ID NO:370.
 XX
 KW Human; secreted protein; cancer; tumour; hyperproliferative disorder;
 KW autoimmune disorder; inflammation; angiogenic diseases; AIDS;
 KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
 KW drug screening; chromosome identification; chromosome mapping;
 KW cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;
 KW antineoplastic; vulnerability; chromosome 9p21; gene; 88.
 XX
 OS Homo sapiens.
 XX
 PN WO200277013-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 26-MAR-2002; 2002WO-US009370.
 XX
 PR 27-MAR-2001; 2001US-0278650P.
 PR 12-SEP-2001; 2001US-00950082.
 PR 12-SEP-2001; 2001US-00950083.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI, 2003-040578/03.
 DR P-PSDB; ABR01306.
 XX
 PT New human secreted proteins and nucleic acids, useful for detecting or
 PT treating cancer or other hyperproliferative disorders, autoimmune
 PT disorders, inflammatory disorders, HIV disease, hepatitis or anemia.
 XX
 PS Claim 21; Page 1345; 2474dp; English.
 XX
 CC ABZ73281-ABZ73697 represent cDNAs corresponding to 391 human secreted
 CC protein genes, and ABP01947-ABP01363 represent the proteins they encode.
 CC ABZ73698-ABZ74687 represent human secreted protein genomic fragments. The
 CC invention also encompasses antibodies specific for the secreted proteins,
 CC the use of the secreted proteins in drug screening and recombinant
 CC vectors and host cells comprising a nucleic acid of the invention. The
 CC secreted proteins are thought to be involved in biological activities
 CC associated with cellular signalling, cellular differentiation, cell
 CC migration, proinflammatory activation and neurotransmitter activity. The
 CC secreted proteins, nucleic acids encoding them, antibodies or antibody
 CC fragments specific for the secreted proteins, and modulators of protein
 CC activity are useful for diagnosing or treating cancers or other
 CC hyperproliferative disorders. Additionally, the secreted proteins and
 CC their nucleic acids may also be used in the treatment of autoimmune
 CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
 CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
 CC wound healing. Nucleic acids of the invention may be used for chromosome
 CC identification, chromosome mapping, in gene therapy, for identifying
 CC individuals from minute biological samples, as hybridisation probes, and
 CC as molecular weight markers. The present sequence represents a human
 CC secreted protein-encoding cDNA clone of the invention
 XX
 SO Sequence 1460 BP; 292 A; 472 C; 398 G; 298 T; 0 U; 0 Other;
 Query Match 96.7%; Score 1392.8; DB 7; Length 1460;
 Best Local Similarity 99.3%; Pred. No. 7.5e-301;
 Matches 1430; Conservative 7; Indels 3; Gaps 3;
 QY 1 AAAGTAACGGCTACAGACAGTGAAGTAATGTTGCTGCGCGGCTAGAAAACTCTGTGCG 60
 DB 22 AAAGTAACGGCTACAGACAGTGAAGTAATGTTGCTGCGCGGCTAGAAAACTCTGTGCG 81
 QY 61 GTACCAACCCAGAGCGTTGAGACAGCCCACTTCAAGGCTTAAACGAGAGGTGCA 120
 DB 82 GTACCAACCCAGAGCGTTGAGACAGCCCACTTCAAGGCTTAAACGAGAGGTGCA 141

QY	121	GGAGCTCAGACTTCAACAGACCCACTGCGTCCAGGCTTGTAACGGAAGAAAGACCCCAAGAC	180
Dp	142	GGAGCTCAGACTTCAACAGACCCACTGCGTCCAGGCTTGTAACGGAAGAAAGACCTCAAGAC	201
QY	181	GGGCTTCTCCCGAGTCCAGAGCAACCCCAAGTTGCTGGGTCGCTGCGCGCTGCGTGCAGC	240
Dp	202	GCGCTCTCCCGAGTCCAGAGCAACCCCAAGTTGCTGGGTCGCTGCGCGCGCTGCGTGCAGC	261
QY	241	ACTCGGCGCGGCTGACGATACCTGTGGAAACGCGGTACTGGCTTTTAAACCCAGGCC	300
Dp	262	ACTCGGCGCGGCTGACGATACCTGTGGAAACGCGGTACTGGCTTTTAAACCCAGGCC	321
QY	301	CGGCATGCGCGAGCTTCAAGGTTCCACTGCTCATCGTATTCTTAAGTGTTTTGGCTCTA	360
Dp	322	CGGCATGCGCGAGCTTCAAGGTTCCACTGCTCATCGTATTCTTAAGTGTTTTGGCTCTA	381
QY	361	GCAGCAAGCTTCTGCTCATCTTGTCGCGGGGATCCGTGGACACTCGGCTGTTTGATGTG	420
Dp	382	GCAGCAAGCTTCTGCTCATCTTGTCGCGGGGATCCGTGGACACTCGGCTGTTTGATGTG	441
QY	421	GTAAGAGTTCTTCTCAGTCTGTTCAATAGCGCGAATAATGTGGCTGTGCATTCAGTCA	480
Dp	442	GTAAGAGTTCTTCTCAGTCTGTTCAATAGCGCGAATAATGTGGCTGTGCATTCAGTCA	501
QY	481	GAATGATTTGTTGGGATACAGTAACACCAACATCTCAAAACCTTCAGCGCACGCGC	540
Dp	502	GAATGATTTGTTGGGATACAGTAACACCAACATCTCAAAACCTTCAGCGCACGCGC	561
QY	541	GTTTACAGCCCGTGGTCTGCTGTGGGCTTGGAGGGCATTAATTAATCACTCACAGG	600
Dp	562	GTTTACAGCCCGTGGTCTGCTGTGGGCTTGGAGGGCATTAATTAATCACTCACAGG	621
QY	601	ACCCGAGTCATCAGCTGATGACAGACCAATGACTACAAACAGCAATTCACCTGGGCTTG	660
Dp	622	ACCCGAGTCATCAGCTGATGACAGACCAATGACTACAAACAGCAATTCACCTGGGCTTG	681
QY	661	AAAGAAATTAACGCCCGCGAGTAACCGCAACGCATCGAAGAAAGGGGCTGCCGACCCAGT	720
Dp	682	AAAGAAATTAACGCCCGCGAGTAACCGCAACGCATCGAAGAAAGGGGCTGCCGACCCAGT	741
QY	721	CTCTACCTGCGAGAAAGTTCAACACGAGTAGCCCTTGCGGCTGTACCAACAGTACAC	780
Dp	742	CTCTACCTGCGAGAAAGTTCAACACGAGTAGCCCTTGCGGCTGTACCAACAGTACAC	801
QY	781	CTGGGCGGACATACGCCCTCGGCGACGTAATGGGTGGCATTCTGCTTCTGCTCTCTCC	840
Dp	802	CTGGGCGGACATACGCCCTCGGCGACGTAATGGGTGGCATTCTGCTTCTGCTCTCTCC	861
QY	841	AAAGTCTCTCTCCACGCGCGGCGCGCTTACAGAAAGCTTGCACTGTGACACACGGA	900
Dp	862	AAAGTCTCTCTCCACGCGCGGCGCGCTTACAGAAAGCTTGCACTGTGACACACGGA	921
QY	901	GCGTGGGCGCTTTCGGGGCTTTCGGGCTTTCGGGCTTTCGATCTTCGAGGTGCGCTGCGCG	960
Dp	922	GCGTGGGCGCTTTCGGGGCTTTCGGGCTTTCGGGCTTTCGATCTTCGAGGTGCGCTGCGCG	981
QY	961	CTCGGCTTAGGCTCTCCGCGCTCAACATAGTACGCGCGCTTCTTGGGTACAGCTG	1020
Dp	982	CTCGGCTTAGGCTCTCCGCGCTCAACATAGTACGCGCGCTTCTTGGGTACAGCTG	1041
QY	1021	GCACCGGAGTCTGTGGCTCTTCTCGGAGAGGGCGGTGAGTCTCCAGTATGTTGCG	1080
Dp	1042	GCACCGGAGTCTGTGGCTCTTCTCGGAGAGGGCGGTGAGTCTCCAGTATGTTGCG	1101
QY	1081	CCGACGCGCTTTGCAACCCCTTGAGACCAAGCGCAAGATGTC-AGCGAGAGAGAG	1139
Dp	1102	CCGACGCGCTTTGCAACCCCTTGAGACCAAGCGCAAGATGTC-AGCGAGAGAGAG	1161
QY	1140	GGGCTACCGCTTATCTGCGGAGACCGCATGACCAAGCAGCGGCTCTCCAGAC-TTAA	1198
Dp	1162	GGGCTACCGCTTATCTCT-GGGGAACCATGACAAAGAGCGCGCTTCCCAACTTTAA	1220

Oy		AANGATACACCCTAACCTGTGAAGGGGAGCCCAATCTGAACTCCTCCCCGCTTGGGAC	1258			
Oy	1199	AANGATACACCCTAACCTGTGAAGGGGAGCCCAATCTGAACTCCTCCCCGCTTGGGAC	1258			
Dd	1221	AANTGATACACCACTTAACCTGTGAAGGGGAGCCCAATCTGAACTCCTCCCCGCTTGGGAC	1280			
Oy	1259	ATTCGACAGGCCCGGGAAGCACGTGTCGCCGCACAGAGAGACTCCAGAAGAGGACAC	1318			
Dd	1281	ATTGCAAGGCCCGGGAAGCACGTGTCGCCGCACAGAGAGACTCCAGAAGAGGACAC	1340			
Oy	1319	TGAGCGCTGCTGGCCGACAGGCGCTTGAGCATTCGCGACGACAAGGAAAAGTTCTCTGCGAGC	1378			
Dd	1341	TGAGCGCTGCTGGCCGACAGGCGCTTGAGCATTCGCGACGACAAGGAAAAGTTCTCTGCGAGC	1400			
Oy	1379	GATCTGTAAATAACCTTTTTTCTTGTGTTTTTAAAAAAGAAAAAAAAAAAAAAAAA	1438			
Dd	1401	GATCTGTAAATAACCTTTTTTCTTGTGTTTTTAAAAAAGAAAAAAAAAAAAAAAAA	1460			
<hr/>						
RESULT 8						
ID	ADDC20289	standard; DNA; 1460 BP.				
XX	ADDC20289;					
XX	18-DEC-2003	(first entry)				
DT						
XX	Human secreted protein coding sequence #228.					
DE						
XX						
KM	gene therapy; human; secreted protein; haemopoietic disorder;					
KM	haematological disorder; anaemia; hemophilia; inflammatory disorder;					
KM	inflammatory bowel disease; Crohn's disease; neoplastic disease; cancer;					
KW	leukaemia; wound healing; epithelial cell proliferation disorder;					
KV	immune disorder; autoimmune disorder; asthmatic disorder;					
KV	cardiovascular disorder; atherosclerosis; myocarditis;					
KX	infectious disease; HIV; AIDS; endocrine disorder; diabetes;					
KX	gastrointestinal disorder; duodenal ulcer; gastroenteritis; gene; de.					
OS	Homo sapiens.					
PN	WO200292787-A2.					
PD	21-NOV-2002.					
PF	26-MAR-2002; 2002WC-US009257.					
PR	27-MAR-2001; 2001US-027865OP.					
PR	12-SEP-2001; 2001US-009500S2.					
PR	12-SEP-2001; 2001US-009500S3.					
PA	(HUMA-) HUMAN GENOME SCI INC.					
PI	Rosen CA,	Ruben SM,				
DE	WPI; 2003-129287/12.					
PT	New human secreted proteins and nucleic acid molecules, useful for					
PT	preparing a diagnostic or pharmaceutical composition for diagnosing,					
PT	preventing or treating hematopoietic or hematologic disorders, e.g.					
PS	anemia or hemophilia.					
Claim 1;	SEQ ID NO 238; 1512bp; English.					
<hr/>						
The invention comprises the amino acid and coting sequences of human						
secreted proteins. The DNA and protein sequences of the invention are						
useful for detecting, preventing, diagnosing, prognosticating, creating						
or ameliorating: haematopoietic or haematological disorders (e.g. anaemia						
and leukaemia); inflammatory disorders (e.g. inflammatory bowel disease						
wound healing and disorders of epithelial cell proliferation); immune						
disorders (e.g. autoimmune disorders and asthmatic disorders);						
cardiovascular disorders (e.g. atherosclerosis and myocarditis);						
infectious disease (e.g. HIV/AIDS); endocrine disorders (e.g. diabetes);						
and gastrointestinal disorders (e.g. duodenal ulcers and						
gastroenteritis). The present DNA Sequence encodes a human secreted						

CC protein of the invention.

Sequence 1460 BP; 292 A; 472 C; 398 G; 298 T; 0 U; 0 Other;

Query Match	Score	DB	Length
96.7%	1392.8	9	1460

Matches 1430; Conservative 0; Mismatches 7; Indels 3; Gaps 3;

Matches 1430; Conservative 0; Mismatches 7; Indels 3; Gaps 3;

QY	AAAGTAAACCGGTACAGACAGTGAAGAAATAGTTCCTGCTCGCGGCTATGAAAACTCTGTCTG	60
Db	AAAGTAAACCGGTAAAGACAGTGAAGAAATAGTTCCTGCTCGCGGCTATGAAAACTCTGTCTG	60
QY	22 AAAGTAAACCGGTAAAGACAGTGAAGAAATAGTTCCTGCTCGCGGCTATGAAAACTCTGTCTG	81
QY	61 GTACCAACCCCAAGCGCTTGAGACACCCCACTCCACGGTTCCTTAAACGAGAGGTGCA	120
Db	82 GTACCAACCCCAAGCGCTTGAGACACCCCACTCCACGGTTCCTTAAACGAGAGGTGCA	141
QY	121 GGACTCAGACTTCAACAGCCCACTCGGTCCAGCCTTGTAACGAAAGAGACGCGCAAGAC	180
Db	142 GGACTCAGACTTCAACAGCCCACTCGGTCCAGCCTTGTAACGAAAGAGACGCTCAAGAC	201
QY	181 GCGCTCTCCCGGTTCAGGCGAGCCCAAGCTTGTCTGCTTGCGCTGCGGCTGCTGTGACG	240
Db	202 GCGCTCTCCCGGTTCAGGCGAGCCCAAGCTTGTCTGCTTGCGCTGCGGCTGCTGTGACG	261
QY	241 ACTCGGCGGCGGTGAGAGATGACCTGTGGAAACGGCGTACGTGCTTTTAAACCCCAAGCC	300
Db	262 ACTCGGCGGCGGTGAGAGATGACCTGTGGAAACGGCGTACGTGCTTTTAAACCCCAAGCC	321
QY	301 CGGCAATCGCGAGGCTTCAAGCGTTCCACTGCTCATCGTTATTTCAAGTGTTTTGCTCTA	360
Db	322 CGGCAATCGCGAGGCTTCAAGCGTTCCACTGCTCATCGTTATTTCAAGTGTTTTGCTCTA	381
QY	361 GCAGCAAGCTTCCTCTCATCTGTGCGGAGAACGCTGGCCACTCGACGCTGATTTTGCTTG	420
Db	382 GCAGCAAGCTTCCTCTCATCTGTGCGGAGAACGCTGGCCACTCGACGCTGATTTTGCTTG	441
QY	421 GTGAGAGTTCCTTCTCAGTCTGTGTCATAGGCGCAGAAATTGTGCTGTGCACTTCAGTGCA	480
Db	442 GTGAGAGTTCCTTCTCAGTCTGTGTCATAGGCGCAGAAATTGTGCTGTGCACTTCAGTGCA	501
QY	481 GAAATGGTTCGAGGGGTACAGTGAACGCAACATCTTCAAAAGCCTTCAGGCGCGCGC	540
Db	502 GAAATGGTTCGAGGGGTACAGTGAACGCAACATCTTCAAAAGCCTTCAGGCGCGCGC	561
QY	541 GTTACAGCCCGTGTGCGGTCTGCTCGTGGGCGCTGAGGGCACTTAAATTATTACTCACAGG	600
Db	562 GTTACAGCCCGTGTGCGGTCTGCTCGTGGGCGCTGAGGGCACTTAAATTATTACTCACAGG	621
QY	601 ACCCAAGTGCATCACTGTAACGAGAACATTTGACTTCAACGAGCACTTCACTGCGCGTGC	660
Db	622 ACCCAAGTGCATCACTGTAACGAGAACATTTGACTTCAACGAGCACTTCACTGCGCGTGC	681
QY	661 AAAGAGATTACGCGCGGAGTACCGCAACGCACTGGAAGAGGGGCTGCGGACCCAGTG	720
Db	682 AAAGAGATTACGCGCGGAGTACCGCAACGCACTGGAAGAGGGGCTGCGGACCCAGTG	741
QY	721 CTCCTACCTGCGGAGAAATTTCACACCGAGTACCTTGCCTGCTTACCACTAGTACAC	780
Db	742 CTCCTACCTGCGGAGAAATTTCACACCGAGTACCTTGCCTGCTTACCACTAGTACAC	801
QY	781 CTGGGCGGAGCACTACGCTCGGCGCAAGCTATGGGTGGCGTTCGTTCTGGCTCCTCTCC	840
Db	802 CTGGGCGGAGCACTACGCTCGGCGCAAGCTATGGGTGGCGTTCGTTCTGGCTCCTCTCC	861
QY	841 AACGTGCTGCTCTCCACGCGCGGCGCTTACGGAAGGCTTGGCACTGCTGACCAACGGA	900
Db	862 AACGTGCTGCTCTCCACGCGCGGCGCTTACGGAAGGCTTGGCACTGCTGACCAACGGA	921
QY	901 GCCTTGCGGCTCTTGGGGGTCTTGGCTTGGCTGAGCTCACTAGAGTGGCGGCTGCGCG	960
Db	922 GCCTTGCGGCTCTTGGGGGTCTTGGCTTGGCTGAGCTCACTAGAGTGGCGGCTGCGCG	981
QY	961 CTCGCGCTAAGGCTCTCGCGCTCACCACTCAAGTACGCGCGGCTTCTGGGTACGCTG	1020

Db	982	CTCCGCTTAGGCTCTCCGCGCTTCACCACTCACTACGCGCCGCTTCTGGGTCAAGCTG	1041
QY	1021	GCACCGGAGTCTCTGTGCTTCTCTCGAGGGGACGGTGTAGTCTCCAGTAGTTGCG	1080
Db	1042	GCACCGGAGTCTGTGCTTCTCTCCAGGGGCGGTGTAGTCTCCAGTAGTTGCG	1101
QY	1081	CCGAGGCTTTCGCACTTCTTGGACCAAAAGGCCAAGATGCG - AGCCAGAGAGAGG	1139
Db	1102	CCGAGGCTTTCGCACTTCTTGGACCAAAAGGCCAAGATGCGAAGCCAAAGAGGAG	1161
QY	1140	GGGCTCACTCTTATCCCTGGCGACCCACATGCAAGAGCGGCTCTCCAGAC - TTA	1198
Db	1162	GGGCTCACTCTTATCT - GGGGACCCACTGCAAGAGCGGCTTCTCCAGACTTTAA	1220
QY	1199	AATGTATCACACTAACTGTGAAGGGGACCCAACTGTGACTCTTCCCGCTTGGAC	1258
Db	1221	AATGTATCACACTAACTGTGAAGGGGACCCAACTGTGACTCTTCCCGCTTGGAC	1280
QY	1259	ATCGCAGGCCGGGAAGAGTCCCGCAGGCTTGGCGCAGAGGCTCCAGAGGGCAC	1318
Db	1281	ATCGCAGGCCGGGAAGAGTCCCGCAGGCTTGGCGCAGAGGCTCCAGAGGGCAC	1340
QY	1319	TGAGCGCTCTGGCGCAGAGGCTTGGACATCCGCAAGCACAGGAAATCTTCTGGGGC	1378
Db	1341	TGAGCGCTCTGGCGCAGAGGCTTGGACATCCGCAAGCACAGGAAATCTTCTGGGGC	1400
QY	1379	GACTGTAAATAAAGCTTTTCTTTGTTTTTAAAAAAGGAAAAAAAAAAAAAAAAA	1438
Db	1401	GACTGTAAATAAAGCTTTTCTTTGTTTTTAAAAAAGGAAAAAAAAAAAAAAAAA	1460

RESULT 9
A2A04634

ID AAA94624 standard; DNA; 498 BP

AC AAA94624;

DT 11-JAN-2001 (First entry)

Human CASB618 EST.

KW Epitope; human; CASB618; chromosome 15; tumour; vaccine; cancer; ovarian;
KW colon; autoimmune disease; HLA_A0201; expressed sequence tag; EST; ss.
KW

OS Homo sapiens.

PN WO2000053748-A2

PD 14-SEP-2000.

PF 09-MAR-2000; 2000MO-EP002048.

PR 11-MAR-1999; 99GB-00005607.

XX

XX
XX
C
C
C
C
C
C
C
E
7
6
5
4

XX
CD
WBT: 2000 572260/53

XX
XX
Now human case618 no

PT therapeutic treatment of cancers, particularly ovarian or colon cancer,
PT autoimmune diseases and related conditions.

PS Claim 32; Page 62; 76pp; English.

CC The present sequence is an expressed sequence tag (EST) for human CASB618
CC protein. The gene for human CASB618 is thought to be located on
CC chromosome 15. CASB618 protein and epitopes of CASB618 protein (see
CC AAB6337 to AAB6339) are useful in diagnosing the occurrence of tumour
CC cells and in vaccines for prophylactic and therapeutic treatment of

XX 12-MAY-2003 (first entry)
 DE Secreted protein gene 360 genomic fragment HUFCLJ1, SEQ ID NO:11734.
 XX
 XX
 XX
 KW Human; secreted protein; cancer; tumour; hyperproliferative disorder;
 KW autoimmune disorder; inflammation; angiogenic diseases; AIDS;
 KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
 KW drug screening; chromosome identification; chromosome mapping;
 KW cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;
 KW antineoplastic; vulnary; chromosome 9p21; gene; ds.
 XX
 OS Homo sapiens.
 XX
 XX
 PN MO200277013-A2.
 XX
 XX
 PD 03-OCT-2002.
 XX
 XX
 PF 26-MAR-2002; 2002MO-US009370.
 XX
 XX
 PR 27-MAR-2001; 2001US-0278650P.
 XX
 PR 12-SEP-2001; 2001US-00950082.
 XX
 PR 12-SEP-2001; 2001US-00950083.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX
 PI Rosen CA, Ruben SM;
 XX
 XX
 DR WPI; 2003-040578/03.
 XX
 XX
 PT New human secreted proteins and nucleic acids, useful for detecting or
 PT treating cancer or other hyperproliferative disorder, autoimmune
 PT disorders, inflammatory disorders, HIV disease, hepatitis or anemia.
 XX
 XX
 PS Disclosure; Page 2315-2316; 2474pp; English.
 XX
 XX
 CC AB273281-AB273697 represent cDNAs corresponding to 391 human secreted
 CC protein genes and ABP00947-ABP01363 represent the proteins they encode.
 CC AB273696-AB274687 represent human secreted protein genomic fragments. The
 CC invention also encompasses antibodies specific for the secreted proteins,
 CC the use of the secreted proteins in drug screening and recombinant
 CC vectors and host cells comprising a nucleic acid of the invention. The
 CC secreted proteins are thought to be involved in biological activities
 CC associated with cellular signalling, cellular differentiation, cell
 CC migration, prohormone activation and neurotransmitter activity. The
 CC secreted proteins, nucleic acids encoding them, antibodies or antibody
 CC fragments specific for the secreted proteins, and modulators of protein
 CC activity are useful for diagnosing or treating cancers or other
 CC hyperproliferative disorders. Additionally, the secreted proteins and
 CC their nucleic acids may also be used in the treatment of autoimmune
 CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
 CC (acquired immunodeficiency syndrome), hepatitis, anaemia and to promote
 CC wound healing. Nucleic acids of the invention may be used for chromosome
 CC identification, chromosome mapping, in gene therapy, for identifying
 CC individuals from minute biological samples, as hybridisation probes, and
 CC as molecular weight markers. The present sequence represents a human
 CC secreted protein genomic fragment referred to in the disclosure of the
 CC invention
 XX
 XX
 SO Sequence 5033 BP; 1061 A; 1409 C; 1465 G; 1098 T; 0 U; 0 Other;
 Query Match 28.1%; Score 405.4; DB 7; Length 5033;
 Best Local Similarity 98.6%; Pred. No. 2e-80;
 Matches 409; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 121 GAACTCAGACTTACACGAGCCACTCGTCCAGCCCTTGTACCAAGAGACGCCAAGAC 180
 DB 133 GAACTCAGACTTACACGAGCCACTCGTCCAGCCCTTGTACCAAGAGAGCCAGAC 192
 QY 181 GCGCTTCCCGCGCTCCAGGAGCCCAAGCTTGTGCTTGCCTGCCCCGCGGAGC 240
 DB 193 GCGCTTCCCGCGCTCCAGGAGCCCAAGCTTGTGCTTGCCTGCCCCGCGGAGC 252
 QY 241 ACTCGGCGGCGTGCAGCATGATACCTGTGGAAGCGGCGTACTGCTTTTACCCAGGCC 300
 DB 253 ACTCGGCGGCGTGCAGCATGATACCTGTGGAAGCGGCGTACTGCTTTTACCCAGGCC 312
 QY 301 CGGATCCCGGAGGCTTACGCTTCCACTGCTCATGTATTTAGTGTGCTCTA 360
 DB 313 CGGATCCCGGAGGCTTACGCTTCCACTGCTCATGTATTTAGTGTGCTCTA 372
 QY 361 GCAGCAAGCTTCCGCTCATCTTGCAGGAGGATCCGAGGCACATCGCGCTGTTT 415
 DB 373 GCAGCAAGCTTCCGCTCATCTTGCAGGAGGATCCGAGGCACATCGGTAGGATGT 427

RESULT 12

ADCC21005
 ID ADCC21005 standard; DNA; 5033 BP.

AC ADCC21005;

DT 18-DEC-2003 (first entry)

DE Human secreted protein-related DNA sequence #423.

XX gene therapy; human; secreted protein; haemopoietic disorder;
 KW haematological disorder; anaemia; haemophilia; inflammatory disorder;
 KW inflammatory bowel disease; Crohn's disease; neoplastic disease; cancer;
 KW leukaemia; wound healing; epithelial cell proliferation disorder;
 KW immune disorder; autoimmune disorder; asthmatic disorder;
 KW cardiovascular disorder; atherosclerosis; myocarditis;
 KW infectious disease; HIV; AIDS; endocrine disorder; diabetes;
 KW gastrointestinal disorder; duodenal ulcer; gastroenteritis; gene; ds.
 XX
 OS Homo sapiens.
 XX
 XX
 PN WO200232787-A2.
 XX
 XX
 PD 21-NOV-2002.
 XX
 XX
 PF 26-MAR-2002; 2002MO-US009257.
 XX
 XX
 PR 27-MAR-2001; 2001US-0278650P.
 XX
 PR 12-SEP-2001; 2001US-00950082.
 XX
 PR 12-SEP-2001; 2001US-00950083.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX
 PI Rosen CA, Ruben SM;
 XX
 XX
 DR WPI; 2003-129287/12.
 XX
 XX
 PT New human secreted proteins and nucleic acid molecules, useful for
 PT preparing a diagnostic or pharmaceutical composition for diagnosing,
 PT preventing or treating hematopoietic or hematologic disorders, e.g.
 PT anemia or hemophilia.
 XX
 XX
 PS Disclosure; SEQ ID NO 959; 1512pp; English.
 XX
 XX
 CC The invention comprises the amino acid and coding sequences of human
 CC secreted proteins. The DNA and protein sequences of the invention are
 CC useful for detecting, preventing, diagnosing, prognosticating, treating
 CC or ameliorating: haematopoietic or haematological disorders (e.g. anaemia
 CC and haemophilia); inflammatory disorders (e.g. inflammatory bowel disease
 CC and Crohn's disease); neoplastic disease (e.g. cancer and leukaemia);
 CC wound healing and disorders of epithelial cell proliferation; immune
 CC disorders (e.g. autoimmune disorders and asthmatic disorders);

CC cardiovascular disorders (e.g. atherosclerosis and myocarditis);
CC infectious disease (e.g. HIV/AIDS); endocrine disorders (e.g. diabetes);
CC and gastrointestinal disorders (e.g. duodenal ulcers and
CC gastroenteritis). The present DNA sequence was used in the
CC exemplification of the invention.

SQ Sequence 5033 BP; 1061 A; 1409 C; 1465 G; 1098 T; 0 U; 0 Other;

Query Match 28.1%; Score 405.4; DB 9; Length 5033;
Best Local Similarity 98.6%; Pred. No. 2e-80; Indels 0; Gaps 0;
Matches 409; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 AAAGTAAACGGCTACAGACAGTGAAGAAATAGTTGCTGCGCGGCTAGAAAAAAGCTGTGCG 60
Db 13 AAAGTAAACGGCTACAGACAGTGAAGAAATAGTTGCTGCGCGGCTAGAAAAAAGCTGTGCG 72
QY 61 GTACCAACCCCAAGGCTTGAAG 120
Db 73 GTACCAACCCCAAGGCTTGAAG 132
QY 121 GAAGTAAACGGCTACAGACAGTGAAGAAATAGTTGCTGCGCGGCTAGAAAAAAGCTGTGCG 180
Db 133 GAAGTAAACGGCTACAGACAGTGAAGAAATAGTTGCTGCGCGGCTAGAAAAAAGCTGTGCG 192
QY 181 GCGCTCTCCGCGCTGACAG 240
Db 193 GCGCTCTCCGCGCTGACAG 252
QY 241 ACTGCGCGCGGCTGACAG 300
Db 253 ACTGCGCGCGGCTGACAG 312
QY 301 CGGATGCGCGAGGCTTCAAG 360
Db 313 CGGATGCGCGAGGCTTCAAG 372
QY 361 GCGATGCGCGAGGCTTCAAG 420
Db 373 GCGATGCGCGAGGCTTCAAG 427

RESULT 13
ADD19187
ID ADD19187 standard; cDNA: 580 BP.

AC ADD19187;
XX
DT 15-JAN-2004 (first entry)
XX
DE Human cDNA from secreted protein gene 4.

XX human secreted protein; cytosolic; antibacterial; vinorelbine;
XX neuroprotective; gynaecological; gastrointestinal; cardiac;
XX cardiovascular; nephrotoxic; anti-inflammatory; muscular;
XX respiratory; immunosuppressive; cerebroprotective; vasodilator;
XX nociceptive; antiallergic; cancer; bacterial infection; viral infection;
XX muscular disorder; immune system disorder; blood disorder;
XX pulmonary disorder; reproductive disorder; gastrointestinal disorder;
XX inflammatory disorder; proliferative disorder; renal disorder;
XX gene.

OS Homo sapiens.
XX
XX WO2003052377-A2.
XX
XX 26-JUN-2003.
XX
XX 06-NOV-2002; 2002WO-US035606.
XX
XX 07-NOV-2001; 2001US-0331046P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX

PI Rosen CA, Ruben SM;
XX
XX WPI: 2003-533050/50.
DR P-PSDB; ADD19262.
XX

PT New isolated nucleic acids encoding signal transduction pathway component
PT polypeptides, useful for diagnosing, treating, and/or preventing
PT disorders, such as cancer, infections, cardiovascular and inflammatory
PT diseases.

PS Claim 1, SEQ ID NO 14; 554pp; English.

XX The invention relates to an isolated nucleic acid molecule (cDNA)
XX encoding a human secreted protein, representing one of 85 novel genes.
XX Also included are recombinant vectors, host cells (expressing the
XX protein), the secreted proteins (including their fragments, epitopes and
XX homologues), an isolated antibody that binds specifically to the protein,
XX diagnosing a pathological condition or susceptibility to a pathological
XX condition (comprising determining the presence or absence of a mutation
XX in the nucleic acid and diagnosing a condition based on the presence or
XX absence of the mutation), diagnosing a pathological condition or
XX susceptibility to a pathological condition (comprising determining the
XX presence or amount of expression of the protein in a biological sample
XX and diagnosing a condition based on the presence or amount of expression
XX of the protein), preventing, treating or ameliorating a medical condition
XX by administering the nucleic acid or protein to a mammalian subject,
XX identifying a binding partner to the protein, the gene corresponding to
XX the cDNA sequence, and identifying an activity in a biological assay
XX (comprising expressing the nucleic acid in a cell, isolating the
XX supernatant, detecting an activity in a biological assay and identifying
XX the protein in the supernatant having the activity). The nucleic acids
XX and proteins display the following activities: Cytostatic, antibacterial,
XX Vinorelbine, Neuroprotective, Gynaecological, Gastrointestinal-Gen,
XX Cardiac, Cardiovascular-Gen, Nephrotoxic, Cerebroprotective, Vasodilator-
XX Gen, Respiratory-Gen, Immunosuppressive, Cerebroprotective, Vasodilator-
XX Gen, Antiallergic. The methods and compositions of the present
XX invention are useful for diagnosing, treating, preventing and/or
XX prognosticating disorders related to the novel polypeptides, such as
XX cancer, bacterial or viral infections, and neural, immune system, blood,
XX muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
XX renal, inflammatory or proliferative disorders (many examples of these
XX diseases and disorders are given in the specification). The present
XX sequence encodes a novel secreted protein of the invention.

SQ Sequence 580 BP; 115 A; 198 C; 156 G; 111 T; 0 U; 0 Other;

Query Match 28.1%; Score 405; DB 9; Length 580;
Best Local Similarity 100.0%; Pred. No. 1.4e-80; Indels 0; Gaps 0;
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAGTAAACGGCTACAGACAGTGAAGAAATAGTTGCTGCGCGGCTAGAAAAAAGCTGTGCG 60
Db 13 AAAGTAAACGGCTACAGACAGTGAAGAAATAGTTGCTGCGCGGCTAGAAAAAAGCTGTGCG 72
QY 61 GTACCAACCCCAAGGCTTGAAG 120
Db 73 GTACCAACCCCAAGGCTTGAAG 132
QY 121 GAAGTAAACGGCTACAGACAGTGAAGAAATAGTTGCTGCGCGGCTAGAAAAAAGCTGTGCG 180
Db 133 GAAGTAAACGGCTACAGACAGTGAAGAAATAGTTGCTGCGCGGCTAGAAAAAAGCTGTGCG 192
QY 181 GCGCTCTCCGCGCTGACAG 240
Db 193 GCGCTCTCCGCGCTGACAG 252
QY 241 ACTGCGCGCGGCTGACAG 300
Db 253 ACTGCGCGCGGCTGACAG 312
QY 301 CGGATGCGCGAGGCTTCAAG 360
Db 313 CGGATGCGCGAGGCTTCAAG 372

QY 361 GCAGCAAGCTTCTGCTCATCTTGGCCGGGATCCGTGCGCACTCG 405
 DB 373 GCAGCAAGCTTCTGCTCATCTTGGCCGGGATCCGTGCGCACTCG 417

RESULT 14

AA566563/c
 ID AA566563 standard; cDNA; 406 BP.

AA566563;

13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #2367.

KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US008631.

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

PA (HISE-) HISEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR P-PSDB; ABG02376.

PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostic, forensic, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.

PS Claim 1; SEQ ID NO 2367; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostic as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensic, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AA564197-AA594564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

SO Sequence 406 BP; 83 A; 107 C; 138 G; 78 T; 0 U; 0 Other;

Query Match 27.9%; Score 401.8; DB 5; Length 406;

Best Local Similarity 99.5%; Pred. No. 6,4e-80;

Matches 403; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGTAACGGCTACAGACAGTGAATAATGTTCCCTGCGCGCTAGAAAACACTGTGCG 60

DB 405 AAGTAACGGCTACAGACAGTGAATAATGTTCCCTGCGCGCTAGAAAACACTGTGCG 346
 QY 61 GTACCAACCCAGAGGTTGAGAGAGCCACACCTCCAGCTTCCTTACGAGAGGTGCA 120
 DB 345 GAGCAACCCAGAGGTTGAGAGAGCCACCTCCAGCTTCCTTACGAGAGGTGCA 286
 QY 121 GGACTCAGACTTACCAAGCCCACTGATCCAGCCTTTAGCAGAAAGAGAGCCAGAGAC 180
 DB 285 GGACTCAGACTTACCAAGCCCACTGATCCAGCCTTTAGCAGAAAGAGAGCCAGAGAC 226
 QY 181 GCGCTTCCCGCGGTCCAGGACACCCAGCTTCGCTGCGCGCGCGGTGACG 240
 DB 225 GCGCTTCCCGCGGTCCAGGACACCCAGCTTCGCTGCGCGCGCGGTGACG 166
 QY 241 ACTGCGCGCGGTGAGCATGACCTGTGGAAGCGCGTACTGCTTTTACCCGAGCC 300
 DB 165 ACTGCGCGCGGTGAGCATGACCTGTGGAAGCGCGTACTGCTTTTACCCGAGCC 106
 QY 301 CGGCAATGCGCGAGCTTCCAGCGTTCCACTGCTCATCTTATTTAGTGTGCTGTA 360
 DB 105 CGGCAATGCGCGAGCTTCCAGCGTTCCACTGCTCATCTTATTTAGTGTGCTGTA 46
 QY 361 GCAGCAAGCTTCTGCTCATCTTGGCCGGGATCCGTGCGCACTCG 405
 DB 45 GCAGCAAGCTTCTGCTCATCTTGGCCGGGATCCGTGCGCACTCG 1

RESULT 15

ADAS3627/c
 ID ADAS3627 standard; cDNA; 2684 BP.

AC ADAS3627;

DT 20-NOV-2003 (first entry)

DE Human coding sequence, SEQ ID 1195.

KM Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
 KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
 KW inflammatory disease; osteoporosis; neurological disease; gene; ss.

OS Homo sapiens.

PN EP1293569-A2.

PD 19-MAR-2003.

PF 21-MAR-2002; 2002EP-00006586.

PR 14-SEP-2001; 2001JP-00328381.

PR 24-JAN-2002; 2002US-0350435P.

PA (HELI-) HELIX RES INST.

PA (REMS-) RES ASSOC BIOTECHNOLOGY.

PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;

DR WPI: 2003-395539/38.

DR P-PSDB; ADAS5266.

PT New polynucleotides encoding full-length polypeptides, e.g. secretory
 PT and/or membrane proteins, useful for developing medicines for diseases in
 PT which the gene is involved, or as target molecules for gene therapy.

PS Claim 1; SEQ ID NO 1195; 205pp; English.

CC The present invention relates to novel human secretory or membrane
 CC proteins (ADAS4072-ADAS5710) and their coding sequences (ADAS2433-
 CC ADAS4071). The coding sequences are useful in the gene therapy of
 CC diseases caused by abnormalities of the proteins, e.g. cancer.

CC inflammatory diseases, osteoporosis or neurological disease.
XX
SQ Sequence 2684 BP; 599 A; 785 C; 684 G; 616 T; 0 U; 0 Other;

Query Match 27.5%; Score 396; DB 7; Length 2684;
Best Local Similarity 97.6%; Pred. No. 2.1e-78;
Matches 402; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
QY 1007 TCTGGGTCAAGCTGGCAACCGGAGTCTGCTGCTTCTCTGGAGGGGCGGTGAGTC 1066
    |||
Db 2363 TTTCGATCCCAACCGCCACAGGCGCTCTGCTCTTCTCTGGAGGGGCGGTGAGTC 2304
    |||

QY 1067 TCCAGTATGTTGGGCCCGGCGCTTTGGCAACCTTTCTGGACCAAGCGGCAAGACTGCA 1126
    |||
Db 2303 TCCAGTATGTTGGGCCCGGCGCTTTGGCAACCTTTCTGGACCAAGCGGCAAGACTGCA 2244
    |||

QY 1127 GCCAGAGAGAGGGGGGCTCACTCTATCTCTGGCGACCCACTGACAAAGAGGCGGCTC 1186
    |||
Db 2243 GCCAGAGAGAGGGGGGCTCACTCTATCTCTGGCGACCCACTGACAAAGAGGCGGCTC 2184
    |||

QY 1187 TCCAGACTTAATAATGATACCACTAACCTGTGAGGGGAGCCCAATCTGACTCTTCC 1246
    |||
Db 2183 TCCAGACTTAATAATGATACCACTAACCTGTGAGGGGAGCCCAATCTGACTCTTCC 2124
    |||

QY 1247 CCGCTTGGGACATCGAGGCCCGGGAAGCAAGTCCCGCAGGCTGGGCGAGAGAGCTC 1306
    |||
Db 2123 CCGCTTGGGACATCGAGGCCCGGGAAGCAAGTCCCGCAGGCTGGGCGAGAGAGCTC 2064
    |||

QY 1307 CAGGAAGGCACTGAGCGCTGTGCGCGAGGCGCTGACATCCGCAAGGCAACGAGGAAA 1366
    |||
Db 2063 CAGGAAGGCACTGAGCGCTGTGCGCGAGGCGCTGACATCCGCAAGGCAACGAGGAAA 2004
    |||

QY 1367 GTCTCTGCGGCGCACTGTAAATAAACCTTTTCTTTGTTTTTAAAAA 1418
    |||
Db 2003 GTCTCTGCGGCGCACTGTAAATAAACCTTTTCTTTGTTTTTAAAAA 1952
    |||
```

Search Completed: February 23, 2004, 19:50:04
Job time : 586.643 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 19:35:58 ; Search time 130.797 Seconds

(without alignments)
6113.919 Million cell updates/sec

Title: US-09-936-456-1

Perfect score: 1441

Sequence: 1 aaagtaacgctacagacag.....aaaaaaaaaaaaaaaa 1441

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 27747546 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PTCUTS.COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	61	4.2	7218	1	US-08-232-463-14
2	50.2	3.5	1098	4	US-09-252-991A-1304
3	50.2	3.5	1257	4	US-09-252-991A-1255
4	50.2	3.5	1746	4	US-09-252-991A-1163
5	48.2	3.3	396	4	US-09-252-991A-1350
6	48	3.3	1436	4	US-09-614-912-35
7	47.6	3.3	420	4	US-09-252-991A-11029
8	47.6	3.3	1683	4	US-09-252-991A-11226
9	47.6	3.3	1974	4	US-09-252-991A-11111
10	47.6	3.3	2283	4	US-09-252-991A-10956
11	46.4	3.2	1431	4	US-09-614-912-49
12	45	3.1	1926	4	US-09-249-585A-4
13	45	3.1	1931	2	US-09-130-114-2
14	44.4	3.1	1835	4	US-09-488-549-1
15	44.4	3.1	15231	3	US-09-128-155-16
16	43.8	3.0	462	4	US-09-252-991A-1610
17	43.8	3.0	1362	4	US-09-252-991A-1556
18	43.8	3.0	1551	4	US-09-252-991A-12931
19	43.8	3.0	1587	4	US-09-252-991A-1561
20	43.8	3.0	1659	4	US-09-252-991A-13081
21	43.8	3.0	1917	4	US-09-252-991A-1465
22	43.8	3.0	2073	4	US-09-252-991A-12858
23	43.8	3.0	5852	1	US-07-867-106-2
24	43.6	3.0	3722	4	US-10-164-595-9
25	43.6	3.0	3862	4	US-10-164-595-5
26	43.6	3.0	3937	4	US-10-164-595-7
27	43.6	3.0	3985	4	US-10-164-595-3

ALIGNMENTS

28	43.4	3.0	597	4	US-09-252-991A-5066	Sequence 5066, Ap
29	43.4	3.0	1308	4	US-09-252-991A-5127	Sequence 5127, Ap
30	43.4	3.0	1782	4	US-09-252-991A-5094	Sequence 5094, Ap
31	43.2	3.0	561	4	US-09-252-991A-13572	Sequence 13572, A
32	43.2	3.0	4131	4	US-09-252-991A-13773	Sequence 13773, A
33	43.2	3.0	8211	4	US-09-252-991A-13656	Sequence 13656, A
34	43	3.0	835	6	5223425-7	Patent No. 5223425
35	43	3.0	1093	6	5223425-3	Patent No. 5223425
36	43	3.0	2089	4	US-09-291-922-25	Sequence 25, Appl
37	43	3.0	2196	4	US-09-673-395A-548	Sequence 548, Appl
38	42.8	3.0	1454	4	US-09-372-422A-19	Sequence 19, Appl
39	42.6	3.0	367	3	US-09-328-111-446	Sequence 446, Appl
40	42.6	3.0	744	4	US-09-252-991A-5564	Sequence 5564, Ap
41	42.6	3.0	1332	4	US-09-252-991A-5561	Sequence 5561, Ap
42	42.6	3.0	1681	4	US-09-434-288-7	Sequence 7, Appl
43	42.4	2.9	68750	3	US-09-335-409-1	Sequence 1, Appl
44	42.4	2.9	68750	4	US-09-568-102-1	Sequence 1, Appl
45	42.4	2.9	68750	4	US-09-567-969-1	Sequence 1, Appl

RESULT 1
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESS: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZ5pc-F1s
US-08-232-463-14
Query Match 4.2%; Score 61; DB 1; Length 7218;

Db 1343 TGGGGCGGGGCTGGGGCCGAACCGGGACCTGCGCTGGCGCTGACACCGGGGCG 1402
QY 950 CGCTCTCCCGCTCCGCTAGAGCTCTCCGCGCTCACCACTAGTACGGCGCGCTTCT 1009
Db 1403 TGTCTGGCGCGCTGTGTCGCCGAAGCCCTGGAGAACCCCGGCGACGCCCGCGAGGCT 1462
QY 1010 GGGTCAGCTGGCAACCGCGGCTCTGTGCTCTCTCTCCGAGAGGGCGCTGTGAGTCTC 1068
Db 1463 TGGCGCTGGCGCGGAGGAGCGCTCGCGGCATTGCTATGACACCTCTGCGGGGAGTC 1521

RESULT 5

US-09-252-991A-1350/C
Sequence 1350 Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 1350
LENGTH: 396
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1350

Query Match 3.3%; Score 48.2; DB 4; Length 396;
Best Local Similarity 52.2%; Pred. No. 0.024;
Matches 107; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 830 GGCTCTCTCCACAGTGTGCTCTCCACCGCGCGCGCTCTACGAGGCTTGACATGCG 889
Db 217 GACTCTGCTCATTAGGGCTGCGCGGCTGCGAGCTGCTGCGGCGCGCTGATGTC 158
QY 890 TGACACCGAGCGCTTCCGCTTTGGGGCTTTGCGCTTCCATCTTACGCTGC 949
Db 157 TGGCGCGCGGCTCGGGCGGAACGCGGAGACCTCGCTGCGCTGACACCGCGCGCG 98
QY 950 CGCTCTCCCGCTCCGCTAGAGCTCTCCGCGCTCACCACTAGTACGGCGCGCTTCT 1009
Db 97 TGTCTGGCGCGCTGTGTCGCCGAAGCCCTGGAGAACCCCGGCGACGCCCGCGAGGCT 38
QY 1010 GGGTCAGCTGGCAACCGCGGCTCT 1034
Db 37 TGGCGCTGGCGCGGAGCGGAGCGCT 13

RESULT 6

US-09-614-912-35
Sequence 35 Application US/09614912
Patent No. 6677502
GENERAL INFORMATION:
APPLICANT: Allen, Steve
APPLICANT: Rafalski, Antoni
APPLICANT: Orozco, Buddy
APPLICANT: Miao, Gou-Hau
APPLICANT: Famodu, Omoayo O.
APPLICANT: Lee, Jian Ming
APPLICANT: Sakai, Hajime
APPLICANT: Meng, Zude
APPLICANT: Cai, Perry G
APPLICANT: Anderson, Shawn
TITLE OF INVENTION: Plant Metabolism Genes
FILE REFERENCE: B01378 US NA
CURRENT APPLICATION NUMBER: US/09/614,912
CURRENT FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: 60/143,401
PRIOR FILING DATE: 1999-07-12
PRIOR APPLICATION NUMBER: 60/143,412
PRIOR FILING DATE: 1999-07-12
PRIOR APPLICATION NUMBER: 60/146,650
PRIOR FILING DATE: 1999-07-30
PRIOR APPLICATION NUMBER: 60/170,906
PRIOR FILING DATE: 1999-12-15
PRIOR APPLICATION NUMBER: 60/172,959
PRIOR FILING DATE: 1999-12-21
PRIOR APPLICATION NUMBER: 60/172,946
NUMBER OF SEQ ID NOS: 204
SOFTWARE: Microsoft Office 97
SEQ ID NO 35
LENGTH: 1436
TYPE: DNA
ORGANISM: Zea mays
US-09-614-912-35

Query Match 3.3%; Score 48; DB 4; Length 1436;
Best Local Similarity 46.8%; Pred. No. 0.039;
Matches 187; Conservative 0; Mismatches 210; Indels 3; Gaps 1;

QY 541 GTTACAGCCCGGTGCTGCTGCTGCTGCGGCGCTGAGGCGCATTAATTACATCAGAGG 600
Db 43 GCTTCT 102
QY 601 ACCCGATGATCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 660
Db 103 TTGACAGCTGCGCGCTGAGCGGCTCTCCGACATCCGAGCATGTTATGCTGCGGCG 162
QY 661 AAGAGATTAACCGCGCGGAGTACCGGAGTACCGGAGTACCGGAGTACCGGAGTAC 720
Db 163 GAGGAGAGCG 222
QY 721 CTCTACCTGGCGGAGAGTTCACAGCG--AGTACCTTGGCGGCTGCTACACAGTAC 777
Db 223 GGGAGCGCGCGCGAGTGTGCGGAGTGTGCGGAGTGTGCGGAGTGTGCGGAGTGTG 282
QY 778 CACTTGGCGGAGACTAGCGCTTGGCGGAGTGTGCGGAGTGTGCGGAGTGTGCGGAG 837
Db 283 CAGGTGTGCGGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATG 342
QY 838 TCCAAAGTGTCTCTCCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 897
Db 343 GCTTCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 402
QY 898 GAGCGCTTGGCGCTCTTGGGAGTCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTT 937
Db 403 TGGGCTAGCGGAGCGCTTACGCGGCGGCTTGGCGCTTGGCGCTTGGCGCTTGGCG 442

RESULT 7

US-09-252-991A-11029/C
Sequence 11029 Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 11029
LENGTH: 420
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-11029

Query Match 3.3%; Score 47.6; DB 4; Length 420;
Best Local Similarity 48.2%; Pred. No. 0.033;
Matches 134; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 785 CGGAGACTACGCTCGGCGCAAGCTAGTGGGCTTCTGCTTCCTCCCAAG 844
DB 382 CGGTGAAAACCTTCTTCATAGCCCTGCTGCTCGACATGACCTGCTG 323
QY 845 TCGTCTCTCCAGCGCGCGCCCGCTCTACGAGCGCTGACCTGACCGAGCC 944
DB 322 CGGTACCGAAGCGCGCGCTGTTCAAGACTGACCGTCTTCTGCGATGCGCCT 253
QY 905 TCGGCTCTTCGCGGCTCTTCGCTTCGCTTCATCTGAGCGTCCGCTTCGCCCTCC 964
DB 262 TCTCCCTGAGCTGCTGCGGACCTTCGCTGCTCGCTCGGCGTCTCACTCGGTGACG 203
QY 965 GCTAGGCTCTCCGCGCTCACTACAGTACGAGCGCGCTTCCTGCTGACGCTGCA 1024
DB 202 CTTGCGCTCGACCGCGAGCGCGGCTTCTGCTTCCTGCTCTGCTGCTGCTG 143
QY 1025 CGGCGCTCTGCTGCTCTCTTCCTCGAGGCGCGCTGCTG 1062
DB 142 GCGGCTCGCTGACCTGTTCCGCTGCGCGCGCGCTG 105

RESULT 8

US-09-252-991A-11226
; Sequence 11226, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11226
; LENGTH: 1683
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11226

Query Match 3.3%; Score 47.6; DB 4; Length 1683;
Best Local Similarity 48.2%; Pred. No. 0.051;
Matches 134; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 785 CGGAGACTACGCTCGGCGCAAGCTAGTGGGCTTCTGCTTCCTCCCAAG 844
DB 864 CGGTGAAAACCTTCTTCATAGCCCTGCTGCTCGACATGACCTGCTG 923
QY 845 TCGTCTCTCCAGCGCGCGCCCGCTCTACGAGCGCTGACCTGACCGAGCCT 944
DB 924 CGGTACCGAAGCGCGCGCTGTTCAAGACTGACCGTCTTCTGCGATGCGCCT 983
QY 905 TCGGCTCTTCGCGGCTCTTCGCTTCGCTTCATCTGAGCGTCCGCTTCGCCCTCC 964
DB 984 TCTCCCTGAGCTGCTGCGGACCTTCGCTGCTCGGCGTCTCACTCGGTGACG 1043
QY 965 GCTAGGCTCTCCGCGCTCACTACAGTACGAGCGCGCTTCCTGCTGACGCTGCA 1024
DB 1044 CTTGCGCTCGACCGCGAGCGCGGCTTCTGCTTCCTGCTGCTGCTGCTG 1103
QY 1025 CGGCGCTCTGCTCTCTTCCTCGAGGCGCGCTGCTG 1062
DB 1104 GCGGCTCGCTGACCTGTTCCGCTGCGCGCGCGCTG 1141

RESULT 9

US-09-252-991A-11111
; Sequence 11111, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONA
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11111
; LENGTH: 1974
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11111

Query Match 3.3%; Score 47.6; DB 4; Length 1974;
Best Local Similarity 48.2%; Pred. No. 0.054;
Matches 134; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 785 CGGAGACTACGCTCGGCGCAAGCTAGTGGGCTTCTGCTTCCTCCCAAG 844
DB 721 CGGTGAAAACCTTCTTCATAGCCCTGCTGCTCGACATGACCTGCTG 790
QY 845 TCGTCTCTCCAGCGCGCGCCCGCTCTACGAGCGCTGACCTGACCGAGCCT 904
DB 791 CGGTACCGAAGCGCGCGCTGTTCAAGACTGACCGTCTTCTGCGATGCGCCT 850
QY 905 TCGGCTCTTCGCGGCTCTTCGCTTCGCTTCATCTGAGCGTCCGCTTCGCCCTCC 964
DB 851 TCTCCCTGAGCTGCTGCGGACCTTCTGCTGCTCGGCGTCACTCGGTGACG 910
QY 965 GCTAGGCTCTCCGCGCTCACTACAGTACGAGCGCGCTTCCTGCTGACGCTGCA 1024
DB 911 CTTGCGCTCGACCGCGAGCGCGGCTTCTGCTGCTCGGCTCTGCTGCTGCTG 970
QY 1025 CGGCGCTCTGCTGCTCTCTTCCTCGAGGCGCGCTGCTG 1062
DB 971 GCGGCTCGCTGACCTGTTCCGCTGCGCGCGCGCTG 1008

RESULT 10

US-09-252-991A-10956/c
; Sequence 10956, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONA
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10956
; LENGTH: 2283
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10956

Query Match 3.3%; Score 47.6; DB 4; Length 2283;
Best Local Similarity 48.2%; Pred. No. 0.056;
Matches 134; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

```

QY 785 CGGAGACATACGCTGCGCCACGCTATAGGATGCGCTTCTGCTCTCTCCACAG 844
Db 2081 CGGTGGAACCGCTTCTTATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2022
QY 845 TGTGCTCTTCCAGCGCGCGCGCTCTACGAGGCTGCTGCTGCTGCTGCTGCTGCT 904
Db 2021 CGGTGACCGAGAGCGCGCGCTGCTTCAAGAGTGAACCGTCTTCTGCGATGCGCGCTT 1962
QY 905 TCGGCTCTTCTGCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 964
Db 1961 TCTGCTGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1902
QY 965 GCTTGGCTCTTCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1024
Db 1901 CTTTGGCTCTTCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1842
QY 1025 CGGAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1062
Db 1841 GCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1804

```

RESULT 11

US-09-614-912-49

Sequence 49, Application US/09614912

Patent No. 6677502

GENERAL INFORMATION:

APPLICANT: Allen, Steve

APPLICANT: Rafalski, Antoni

APPLICANT: Orozco, Buddy

APPLICANT: Miao, Gou-Hau

APPLICANT: Famodu, Omolayo O.

APPLICANT: Lee, Jian Ming

APPLICANT: Sakai, Hajime

APPLICANT: Meng, Zude

APPLICANT: Cai, Perry G

APPLICANT: Anderson, Shawn

TITLE OF INVENTION: Plant Metabolism Genes

FILE REFERENCE: B1378 US NA

CURRENT APPLICATION NUMBER: US/09/614,912

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: 60/143,401

PRIOR FILING DATE: 1999-07-12

PRIOR APPLICATION NUMBER: 60/143,412

PRIOR FILING DATE: 1999-07-12

PRIOR APPLICATION NUMBER: 60/146,650

PRIOR FILING DATE: 1999-07-30

PRIOR APPLICATION NUMBER: 60/170,906

PRIOR FILING DATE: 1999-12-15

PRIOR APPLICATION NUMBER: 60/172,959

PRIOR FILING DATE: 1999-12-21

PRIOR APPLICATION NUMBER: 60/172,946

PRIOR FILING DATE: 1999-12-21

NUMBER OF SEQ ID NOS: 204

SOFTWARE: Microsoft Office 97

SEQ ID NO 49

LENGTH: 1431

TYPE: DNA

ORGANISM: Zea mays

FEATURE:

NAME/KEY: unsure

LOCATION: (1309)

NAME/KEY: unsure

LOCATION: (1339)

NAME/KEY: unsure

LOCATION: (1351)

NAME/KEY: unsure

LOCATION: (1402)

NAME/KEY: unsure

LOCATION: (1429)

```

Best Local Similarity 46.5%; Pred. No. 0.094;
Matches 186; Conservative 0; Mismatches 211; Indels 3; Gaps 1;
QY 541 GTTACAGCCCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
Db 33 GCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 92
QY 601 ACCGAGTGCATCAGCTGACGAGACCACTTGTACATCAACAGAGAGTCACTGAGCTGCTG 660
Db 93 TTGAGAGCTGCGCGCTGAGCGGCTTCCGACATCCCGAGCACTTATGCTGCGCG 152
QY 661 AAGAGATTTAGCGCGCGAGTACCGCAACGACTGAGAAAGAGGCTGCGGACCATG 720
Db 153 GAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 212
QY 721 CTCTAGCTGCGAGAGAGTTCACAGCG---AGTAGCCCTTGCGGCTGCTGCTGCTGCTGCT 777
Db 213 GGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 272
QY 778 CACCTGCGCGGACACTACGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 837
Db 273 CAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 332
QY 838 TCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 897
Db 333 GCTTTTACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 392
QY 898 GAGAGCTTGTGCGCTTCTGCGGCTTCTGCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCT 937
Db 393 TGGGCTACGCGAGAGCTTACAGGCGCGGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 432

```

RESULT 12

US-09-249-585A-4

Sequence 4, Application US/09249585A

Patent No. 6417002

GENERAL INFORMATION:

APPLICANT: Hollick, Robert

TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES

FILE REFERENCE: 0867/00905

CURRENT APPLICATION NUMBER: US/09/249,585A

PRIOR FILING DATE: 1999-02-11

NUMBER OF SEQ ID NOS: 18

SOFTWARE: PatentIn version 3.0

SEQ ID NO 4

LENGTH: 1926

TYPE: DNA

ORGANISM: Epstein Barr Virus

FEATURE:

NAME/KEY: misc.feature

LOCATION: (1)-(1926)

OTHER INFORMATION: template strand of EBNA-1 DNA

US-09-249-585A-4

Query Match

3.1%; Score 45; DB 4; Length 1926;

Best Local Similarity 44.1%; Pred. No. 0.22;

Matches 189; Conservative 0; Mismatches 240; Indels 0; Gaps 0;

```

QY 825 CTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 884
Db 706 CTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 765
QY 885 ACTGTGACCAACCGAGCGCTTGTGCGGCTTGTGCGGCTTGTGCGGCTTGTGCGGCTTGTG 944
Db 766 CTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 825
QY 945 CTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1004
Db 826 CTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 885
QY 1005 CTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1064
Db 886 CTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 945

```

```

Query Match
3.2%; Score 46.4; DB 4; Length 1431;

```

OY	1065	TCTCCAGTAATGTTCCGGCCACGAGGCTCTTCGACCTCTTGAGACCAAGGCGCAAGACTG	1124
Db	946	CCCCGTCCTGCTCTCCGCTCCGCTCTGCTCTTCACCTCCGGCCCCAGCTCTCTGCACCT	1005
OY	1125	CAGCCAGAGAGAGGGGGCTCACTTTATCTCGGCGAACCCACTGCACAGACAGGCGCG	1184
Db	1006	CCGAGCCCAAGCTCTCATCACTCCGGCCCCAGCTCTCTCATCACTCCGGGGGCCCA	1065
OY	1185	TCTCCCACTTAATGTATACACACTAACCTGTGAGGGGAAACCAATCTGACTCCTT	1244
Db	1066	TCTCCCTGCACTTCTTCGGTCCGCCCTTCAGACACTTTCTCGTCCCTCTCCAGACACT	1125
OY	1245	CCCCGACCTT	1253
Db	1126	GCACTCTTT	1134

```

RESULT 13
US-09-130-114-2
: Sequence 2, Application US/09130114
: Patent No. 5976807
:
: GENERAL INFORMATION:
: APPLICANT: Horlick, Robert A.
: APPLICANT: Dama, Bassam B.
: APPLICANT: Robbins, Alan K.
: TITLE OF INVENTION: Eukaryotic Cells Stab-
: TITLE OF INVENTION: From Multiple Transfe-
: FILE REFERENCE: 0867/1D903US1
: CURRENT APPLICATION NUMBER: US/09/130,114
: CURRENT FILING DATE: 1998-08-06
: NUMBER OF SEQ ID NOS: 36
: SOFTWARE: FastSeq for Windows Version 3.0.
:
: SEQ ID NO 2
:
: LENGTH: 1931
: TYPE: DNA
: ORGANISM: EBNA
:
: US-09-130-114-2

```

Query Match	3.1%	Score 45;	DB 2;	Length 1931;
Best Local Similarity	44.1%;	Pred. No. 0.22;		
Matches 189;	Conservative	0;	Mismatches 240;	Indels 0;
			Gaps	0

QY	825	CTTCTGAGCTCTCTCCAAAGTGTGCTCTCCAGCGCGGGCCCGCTCTAGAGAGGCTGGC	884
Db	706	CTGTGTCTCCCGTCTCTGTCTCTCTCCCGTCTCTGTCTCTCTCCCGTCTCTCCGTCTGT	765
QY	885	ACTGTGACCAACCGAGCTTTCGCGCTCTTCGGGGTCTTTCGCTCTTGACCTTCATCTCTAG	944
Db	766	CTTCCCGGTCTCCCGTCTGTGTCTTCCCGTCTCCCGTCTGTGTCTCTCTCCCGT	825
QY	945	CTGTGCGCTTGTGCGCGCTCTCGCTTGGTCTCTCCGCGTCAACACTGATGGAGCGGC	1004
Db	826	CTTCCCGTCTGTGTCTCTCTCCCGTCTCCCGTCTGTGTCTCTCTCCCGTCTCTCCGT	885
QY	1005	CTTCTGGGTCAAGCTGGCAACCGGCGTCTGTGCTCTTCTCTCGAGAGGGGCGTGTGAG	1064
Db	886	CTGTGTCTCCCGTCTCTCCCGTCTGTGTCTCTCTCCCGTCTCTCCGTCTCTCTCT	945
QY	1065	TCTTCAGATGTTGGCGCCAGGCGCTTTCGCAACCTTCTGAGCAAAAGCGCAAGGACTG	1124
Db	946	CCCGGTCTGTCTCTCCCGTCTGTGTCTCTCACTCCGAGCGCCAGGTCTTCGTGACT	1005
QY	1125	CAGCGAGAGAGAGGGGGGTCACTCTTATCTCTGGGAGATCCACTGCAACAAGACGGCG	1184
Db	1006	CGGCGCCAGCTCTTCATCACTTCGAGCCCACTCTTCATCACTCTCGAGGGGCCCA	1065
QY	1185	TCTCCGAGATTAAATGTATCAACCACTAACCTGTGAGGGGAGCCCAATCTGGACTCTT	1244
Db	1066	TCTCTGCACTTCTGTGTGTCTCCCGCTTCAAGACTCTTCTGGTCTCCCGTCTCCAGACT	1125
QY	1245	CCCGGCTT	1253

```

Db          1126 GCACTCTT 1134

RESULT 14
US-09-485-549-1
; Sequence 1, Application US/09485549
; Patent No. 6361948
; GENERAL INFORMATION:
; APPLICANT: James Tricoli
; APPLICANT: Rachel Rhodinelli
; APPLICANT: Fox Chase Cancer Center
; TITLE OF INVENTION: Prognostic Compositions for Prostate Cancer and Methods of U
; TITLE OF INVENTION: thereof
; FILE REFERENCE: PCCC 96-13
; CURRENT APPLICATION NUMBER: US/09/485,549
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: PCT/US98/16768
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: 60/055,285
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1835
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-485-549-1

```

Query Match	3.1%	Score 44.4;	DB 4;	Length 1035;
Best Local Similarity	59.5%;	Pred. No. 0.3;		
Matches	75;	Conservative	0;	Mismatches 51;
			Indels	0;
			Gaps	0;

Qy	131	CACTAGAGGCGCTCTGGCGGAGAGGCGCTGGACATCCGAGGACACAGAGAAAAGTCTCTGG	1375
Db	1658	CACCTTCCGGCCCTGCTGTGGGGGAGAGAGTGAAGGGTTGGGGATCCGGCATGTGGAGG	171
Qy	1376	GGGAGTCTGTAAATAACCTTTTCTTTCTTTTAAATAAAAAAAAAAAAAAAAAAAAAA	1435
Db	1718	GGGGCTCTAGATTAAAGATTACCTTTGGGGTAAATAAAAAAAAAAAAAAAAAAAAAA	1777
Qy	1436	AAAAAA 1441	
Db	1778	AAAAAA 1783	

RESULT 15
US-09-128-155-16

```

: Sequence 16, Application US/09128155
: Patent No. 6117654
: GENERAL INFORMATION:
: APPLICANT: Pat, Yang
: TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
: TITLE OF INVENTION: AND USES THEREOF
: FILE REFERENCE: 09404/052001
: CURRENT APPLICATION NUMBER: US/09/128,155
: EARLIER APPLICATION NUMBER: US 60/091,650
: EARLIER FILING DATE: 1998-07-02
: EARLIER APPLICATION NUMBER: US 60/054,646
: EARLIER FILING DATE: 1997-08-04
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 16
: LENGTH: 152331
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)...(152331)
: OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

```

Query Match	3.18;	Score 44.4;	DB 3;	Length 152331,
-------------	-------	-------------	-------	----------------

Best Local Similarity 51.5%; Pred. No. 1.2;
Matches 102; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY	832	CTCCTTCCAAAGCTGCTCTCCAGCCGCCCCCGCTCTACGAGAGCTGGACTGCTG	891
Db	22095	CCGCCCCCCCCCCCCCGAGCGAGCCCCCCCCCGCCCCCGACCCCCCG	22154
QY	892	ACCAACGAGAGCTTCGCGCTTCGAGGCTTCGCTTCATCTTAGGTCG	951
Db	22155	ACCCCCCGCGCGCGCGCCCCCCCCCGCGCCCCCGCGCGCGCG	22214
QY	952	CTTGGCCGCTCCGCTTAGGCTCTCCGCGCTCACCAGTAGAGCGCGCTTCTGG	1011
Db	22215	CCCCACCCCCCCCCCAGCCCGAGCGCGCCCCCCCCCAGCCCCCGC	22274
QY	1012	GTCAGCTGCGCACCGGC	1029
Db	22275	CCCCGCCCCGAGCCCGC	22292

Search completed: February 23, 2004, 23:23:23
Job time : 132.797 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 19:36:48 ; Search time 552.916 Seconds
(without alignments)
9125.925 Million cell updates/sec

Title: US-09-936-456-1

Perfect score: 1441
Sequence: 1 aaagtaacgctacagacag.....aaaaaaaaaaaaaaaa 1441

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapept 1.0

Searched: 2308684 seqs, 1750822206 residues

Total number of hits satisfying chosen parameters: 4617368

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database:

Published Applications NA:*

- 1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/prodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/prodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/prodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/prodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/prodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/prodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1435.6	99.6	1474	15	US-10-264-237-1097
2	1413	98.1	1421	15	US-10-187-657-2
3	1407.8	97.7	1420	14	US-10-187-657-4
4	694.6	48.2	1594	14	US-10-187-657-9
5	451.6	31.3	522	14	US-10-187-657-5
6	396	27.5	2684	15	US-10-094-749-1195
7	385.8	26.8	450	14	US-10-187-657-7
8	354.8	24.6	1029	10	US-09-759-130B-424
9	354.8	24.6	1029	13	US-10-042-431-54
10	354.8	24.6	2133	13	US-09-759-130B-423
11	354.8	24.6	2133	13	US-10-042-431-53
12	322	22.3	506	14	US-10-187-657-6
13	300	20.8	346	14	US-10-187-657-3
14	245.8	17.1	2684	15	US-10-094-749-1195
15	200.6	13.9	439	15	US-10-027-633-91970

16	196.2	13.6	232	9	US-09-783-590-9492	Sequence 9492, Ap
17	170	11.8	250	14	US-10-187-657-8	Sequence 8, Appl
18	64	4.4	512	14	US-10-106-698-3277	Sequence 3277, Ap
19	50.6	3.5	1455	14	US-10-156-761-5889	Sequence 5889, Ap
20	50.6	3.5	9025608	14	US-10-156-761-1	Sequence 1, Appl
21	50	3.5	594	14	US-10-123-155-10	Sequence 10, Appl
22	50	3.5	594	14	US-10-146-731-10	Sequence 10, Appl
23	50	3.5	594	14	US-10-140-472-10	Sequence 10, Appl
24	50	3.5	594	14	US-10-141-761-10	Sequence 10, Appl
25	50	3.5	594	14	US-10-142-885-10	Sequence 10, Appl
26	50	3.5	594	14	US-10-158-790-10	Sequence 10, Appl
27	50	3.5	594	15	US-10-137-871-10	Sequence 10, Appl
28	50	3.5	594	15	US-10-140-923-10	Sequence 10, Appl
29	50	3.5	594	15	US-10-141-756-10	Sequence 10, Appl
30	50	3.5	594	15	US-10-141-759-10	Sequence 10, Appl
31	50	3.5	594	15	US-10-140-805-10	Sequence 10, Appl
32	50	3.5	594	15	US-10-140-864-10	Sequence 10, Appl
33	50	3.5	1362	14	US-10-156-761-7248	Sequence 7248, Ap
34	50	3.5	9025608	14	US-10-156-761-1	Sequence 1, Appl
35	48.4	3.4	511	12	US-10-424-599-56520	Sequence 56520, A
36	48	3.3	234	10	US-09-814-353-4810	Sequence 4810, Ap
37	48	3.3	234	10	US-09-814-353-11107	Sequence 11107, A
38	48	3.3	380	10	US-09-814-353-17491	Sequence 17491, A
39	48	3.3	1035	15	US-10-260-238-272	Sequence 272, App
40	47.6	3.3	277	9	US-09-960-352-12673	Sequence 12673, A
41	47.4	3.3	263	10	US-09-814-353-5246	Sequence 5246, Ap
42	47.4	3.3	263	10	US-09-814-353-1153	Sequence 1153, A
43	47.2	3.3	434	10	US-09-814-353-17917	Sequence 17917, A
44	46.8	3.2	397	14	US-10-198-846-13054	Sequence 13054, A
45	46.6	3.2	332	10	US-09-814-353-17583	Sequence 17583, A

ALIGNMENTS

RESULT 1
US-10-264-237-1097
; Sequence 1097, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
; APPLICANT: Birste et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PAL13PI
; CURRENT APPLICATION NUMBER: US/10/264,237
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 1097
; LENGTH: 1474
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-264-237-1097

Query Match 99.6%; Score 1435.6; DB 15; Length 1474;

Best local similarity 99.8%; Pred. No. 0;
Matches 1435; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY	1	AAAGTACGGCTTACAGACAGTGAAGAAATGTTTGGCTTGGCGGCTTACGAAATACTGTGCG	60
DB	27	AAAGTACGGCTTACAGACAGTGAAGAAATGTTTGGCTTGGCGGCTTACGAAATACTGTGCG	86
QY	61	GTACCAACCCAGAGGCTTACAGACAGTGAAGAAATGTTTGGCTTGGCGGCTTACGAAATACTGTGCG	120
DB	87	GTACCAACCCAGAGGCTTACAGACAGTGAAGAAATGTTTGGCTTGGCGGCTTACGAAATACTGTGCG	146
QY	121	GAGCTCAGACTTACCAAGCCCACTGCTTCCAGCTTGTTCAGCAAGAGAGCCCAAGAGAC	180
DB	147	GAGCTCAGACTTACCAAGCCCACTGCTTCCAGCTTGTTCAGCAAGAGAGCCCAAGAGAC	206

QY 181 GCGCTCCCGGCTCCAGGAGCCCGAGCTTGGCTTGGCTCCCGGCTGGGCTGAGC 240
 DB 207 GCGCTCCCGGCTCCAGGAGCCCGAGCTTGGCTTGGCTCCCGGCTGGGCTGAGC 266
 QY 241 ACTGGCGGGGCTGAGAGTACCTTGGGAGAGGGCTACTGGCTTTTAAACCCGAGCC 300
 DB 267 ACTGGCGGGGCTGAGAGTACCTTGGGAGAGGGCTACTGGCTTTTAAACCCGAGCC 326
 QY 301 CGGCAATGCGGAGGCTTCACTGCTCATCGTTATTCAGTGTGTTGGCTCA 360
 DB 327 CGGCAATGCGGAGGCTTCACTGCTCATCGTTATTCAGTGTGTTGGCTCA 386
 QY 361 GAGAGAGCTTCTGCTCATCTTGGCGGAGATCCGTGGCCACTGCGGCTGGTTGGTTG 420
 DB 387 GAGAGAGCTTCTGCTCATCTTGGCGGAGATCCGTGGCCACTGCGGCTGGTTGGTTG 446
 QY 421 GTGAGAGTCTTCTGCTCATCTTGGCGGAGATCCGTGGCCACTGCGGCTGGTTGGTTG 480
 DB 447 GTGAGAGTCTTCTGCTCATCTTGGCGGAGATCCGTGGCCACTGCGGCTGGTTGGTTG 506
 QY 481 GATGAGTGTGGGTACAGTAAACAACAACATCTTAACAAGCTTTCAGCGAGCGCC 540
 DB 507 GATGAGTGTGGGTACAGTAAACAACAACATCTTAACAAGCTTTCAGCGAGCGCC 566
 QY 541 GTTACAGCCCGGTGTGGCTGCTGCTGGGGGCTGGAGGGCAATTAATTAACATCAAGGG 600
 DB 567 GTTACAGCCCGGTGTGGCTGCTGCTGGGGGCTGGAGGGCAATTAATTAACATCAAGGG 626
 QY 601 ACCCGAGTGCATCAGCTGAACAGAGAACATTTGACTCAACAGAGCATTCCTGGCTG 660
 DB 627 ACCCGAGTGCATCAGCTGAACAGAGAACATTTGACTCAACAGAGCATTCCTGGCTG 686
 QY 661 AAAGAGATTACGCGCGGAGTACGCGAACAACAATGAGAGAGGGGCTGGAGCCAGAG 720
 DB 687 AAAGAGATTACGCGCGGAGTACGCGAACAACAATGAGAGAGGGGCTGGAGCCAGAG 746
 QY 721 CTCTACTGCGGAGAAAGTTACACAGAGTACGCTTGGCGCTGTATCAACAGAGTACAG 780
 DB 747 CTCTACTGCGGAGAAAGTTACACAGAGTACGCTTGGCGCTGTATCAACAGAGTACAG 806
 QY 781 CTGGCGGAGACTAGAGCTTGGCCAGAGTATGGGTGGCTTGTCTTGGCTCTCTCC 840
 DB 807 CTGGCGGAGACTAGAGCTTGGCCAGAGTATGGGTGGCTTGTCTTGGCTCTCTCC 866
 QY 841 AACGTGCTCTCCAGAGCGGCGCTCTAGAGAGGCTTGGAGTGTGACACAGCGA 900
 DB 867 AACGTGCTCTCCAGAGCGGCGCTCTAGAGAGGCTTGGAGTGTGACACAGCGA 926
 QY 901 GCTTTCGCGCTTCTTGGGAGTCTTGGCTTGGCTTGCATCTTACAGCTGCGCTGCGG 960
 DB 927 GCTTTCGCGCTTCTTGGGAGTCTTGGCTTGGCTTGCATCTTACAGCTGCGCTGCGG 986
 QY 961 CTCCGCTAGAGCTCTCCGCGGCTACCACTGATAGGCGCGCTTCTGGGTCAAGCTG 1020
 DB 987 CTCCGCTAGAGCTCTCCGCGGCTACCACTGATAGGCGCGCTTCTGGGTCAAGCTG 1046
 QY 1021 GCAACCGGAGTCTTGGCTTCTTGGAGAGGCGCTTGGAGTGTGACAGTATGTTCCG 1080
 DB 1047 GCAACCGGAGTCTTGGCTTCTTGGAGAGGCGCTTGGAGTGTGACAGTATGTTCCG 1106
 QY 1081 CCCAGGAGCTTCCAGAGCTTCTGAGAGAGGCGCAAGAGCTGACAGAGAGAGAGG 1140
 DB 1107 CCCAGGAGCTTCCAGAGCTTCTGAGAGAGGCGCAAGAGCTGACAGAGAGAGAGG 1166
 QY 1141 GAGTACCTCTTATCTCGGAGAGCCACTGACAGAGAGGCGGCTTCCAGAGTAAAA 1200
 DB 1167 GAGTACCTCTTATCTCGGAGAGCCACTGACAGAGAGGCGGCTTCCAGAGTAAAA 1226
 QY 1201 TGTATCACACTAAGCTTGAAGAGGAGAGCCATCTGAGCTCTTCCCGGCTTGGAGAT 1260
 DB 1227 TGTATCACACTAAGCTTGAAGAGGAGAGCCATCTGAGCTCTTCCCGGCTTGGAGAT 1286
 QY 1261 CGAGAGCCGGAAGAGTGGCCGAGGAGCTTGGGCGAGAGAGCTTCCAGAGAGGCACTG 1320

DB 1287 CGAGAGCCGGAAGAGAGTGGCCGCGAGGCTGGGCGCAGAGAGCTCCAGAGAGGCACTG 1346
 QY 1321 AGCGCTGCGGCGGAGGCTCGGACATCCGAGAGACAGAGGAAAGTCTCGGGGCGA 1380
 DB 1347 AGCGCTGCGGCGGAGGCTCGGACATCCGAGAGACAGAGGAAAGTCTCGGGGCGA 1406
 QY 1381 TCTGTAATATAACCTTTTCTTTTGTGTTTTTAAAAA 1438
 DB 1407 TCTGTAATATAACCTTTTCTTTTGTGTTTTTAAAAA 1464

RESULT 2

US-10-187-657-2
 ; Sequence 2, Application US/10187657
 ; Publication No. US2003006831A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lasek, Amy K. W.
 ; APPLICANT: Baughn, Mariah R.
 ; APPLICANT: Azimzal, Yalda
 ; TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
 ; FILE REFERENCE: PV-0009 CIP
 ; CURRENT APPLICATION NUMBER: US/10/187,657
 ; CURRENT FILING DATE: 2002-07-01
 ; PRIOR APPLICATION NUMBER: PCT/US00/07817
 ; PRIOR FILING DATE: 2000-03-22
 ; PRIOR APPLICATION NUMBER: 60/139,565
 ; PRIOR FILING DATE: 1999-06-16
 ; NUMBER OF SEQ ID NOS: 9
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 2
 ; LENGTH: 1421
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: Incyte ID No. US2003006831A1 4901066CB1
 US-10-187-657-2

Query Match 98.1%; Score 1413; DB 14; Length 1421;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAGTAAAGGCTACAGAGTGAAGAAATGTTTCGCTCCGCGCTAGAAAACTGTGCG 60
 DB 9 AAAGTAAAGGCTACAGAGTGAAGAAATGTTTCGCTCCGCGCTAGAAAACTGTGCG 68
 QY 61 GTACCAACCCAGAGCTTGAAGAGAGCCCACTTCCAGCTTCTTAACGAGAGGTGCA 120
 DB 69 GTACCAACCCAGAGCTTGAAGAGAGCCCACTTCCAGCTTCTTAACGAGAGGTGCA 128
 QY 121 GGAATCAGACTTACAGAGCCCACTGGGTCCAGCTTGTACGCAAGAGAGCCCAAGAGC 180
 DB 129 GGAATCAGACTTACAGAGCCCACTGGGTCCAGCTTGTACGCAAGAGAGCCCAAGAGC 188
 QY 181 GCGCTTCCCGGCTCCAGGAGAGCCCAAGCTTGTGCTGCTGCTGCGGCTGTGGTACAGC 240
 DB 189 GCGCTTCCCGGCTCCAGGAGAGCCCAAGCTTGTGCTGCTGCTGCGGCTGTGGTACAGC 248
 QY 241 ACTGGGCGGAGTACAGATGACCTGTGGAGAGGAGTACTGCTTTTAAACCCAGAGCC 300
 DB 249 ACTGGGCGGAGTACAGATGACCTGTGGAGAGGAGTACTGCTTTTAAACCCAGAGCC 308
 QY 301 CGGAGTCCGAGGCTTCAAGCTTCACTGCTCATGTTATTTCTAGTGTGTTGGCTCA 360
 DB 309 CGGAGTCCGAGGCTTCAAGCTTCACTGCTCATGTTATTTCTAGTGTGTTGGCTCA 368
 QY 361 GAGAGAGTTCCTGCTCATCTTGGCGGAGAGTCCGAGGAGTCCGAGGAGTGTGTTGGTTG 420
 DB 369 GAGAGAGTTCCTGCTCATCTTGGCGGAGAGTCCGAGGAGTCCGAGGAGTGTGTTGGTTG 428
 QY 421 GTGAGAGTTCCTGCTCATCTTGGCGGAGAGTGTGAGGAGAAATTTGGCTGTGCACTTCACTGCA 480


```

QY 721 CTCTACTGCGGAGAAATTCACACCGAGTAGACCTTGAGGCTGTACCAACAGTACAC 780
DB 729 CTCTACTGCGGAGAAATTCACACCGAGTAGACCTTGAGGCTGTACCAACAGTACAC 788
QY 781 CTGGGCGGAGAAATTCACACCGAGTAGACCTTGAGGCTGTACCAACAGTACAC 840
DB 789 CTGGGCGGAGAAATTCACACCGAGTAGACCTTGAGGCTGTACCAACAGTACAC 848
QY 841 AACGTGCTGCTTCCACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900
DB 849 AACGTGCTGCTTCCACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 908
QY 901 GCGTTGCGGCTCTTCCGAGGCTCTTCCGAGGCTCTTCCGAGGCTCTTCCGAGGCT 960
DB 909 GCGTTGCGGCTCTTCCGAGGCTCTTCCGAGGCTCTTCCGAGGCTCTTCCGAGGCT 968
QY 961 CTCGGCTAGAGCTCTTCCGAGGCTCTTCCGAGGCTCTTCCGAGGCTCTTCCGAGGCT 1020
DB 969 CTCGGCTAGAGCTCTTCCGAGGCTCTTCCGAGGCTCTTCCGAGGCTCTTCCGAGGCT 1028
QY 1021 GCAACCGGCGCTCTTCCGAGGCTCTTCCGAGGCTCTTCCGAGGCTCTTCCGAGGCT 1080
DB 1029 GCAACCGGCGCTCTTCCGAGGCTCTTCCGAGGCTCTTCCGAGGCTCTTCCGAGGCT 1088
QY 1081 CCCAGGCTCTTCCGAGGCTCTTCCGAGGCTCTTCCGAGGCTCTTCCGAGGCTCTTCC 1140
DB 1089 CCCAGGCTCTTCCGAGGCTCTTCCGAGGCTCTTCCGAGGCTCTTCCGAGGCTCTTCC 1148
QY 1141 GCGTCACTCTTATCTCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
DB 1149 GCGTCACTCTTATCTCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1208
QY 1201 TGTATCACCTTATCTCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1260
DB 1209 TGTATCACCTTATCTCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1268
QY 1261 CGCAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1320
DB 1269 CGCAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1328
QY 1321 AGCGTCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380
DB 1329 AGCGTCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1388
QY 1381 TCTGTAAATTAACCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 1441
DB 1389 TCTGTAAATTAACCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 1449

```

RESULT 4

```

US-10-187-657-9
; Sequence 9, Application US/10187657
; Publication No. US20030068311A1
; GENERAL INFORMATION:
; APPLICANT: Laszek, Amy K.W.
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Azimzal, Yalda
; TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
; FILE REFERENCE: PV-0009 CIP
; CURRENT APPLICATION NUMBER: US/10/187,657
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: PCT/US00/07817
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/139,565
; PRIOR FILING DATE: 1999-06-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program
; SEQ ID NO: 9
; LENGTH: 1594
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature

```

OTHER INFORMATION: Inqire ID No. US20030068311A1 110769_Mm.1
US-10-187-657-9

Query Match 48.2%; Score 694.6; DB 14; Length 1594;
Best Local Similarity 75.7%; Pred. No. 3,2e-183;
Matches 918; Conservative 0; Mismatches 279; Indels 16; Gaps 4;

```

QY 222 CTGCGCGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 281
DB 107 CTGCGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 166
QY 282 GCGTTTACCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 341
DB 167 ACCCTTTTACCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 226
QY 342 TCTAGTGTGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 401
DB 227 CTTGATGTTCTTGTCTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 286
QY 402 CTGCGCTGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 461
DB 287 CTGCGCTGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 346
QY 462 GCGTGTGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 521
DB 347 GCGTGTGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 406
QY 522 AGCGTTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 581
DB 407 AGCGTTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 466
QY 582 TAAATTAATCACTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 641
DB 467 TAAATTAATCACTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 526
QY 642 GCGATTACCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 701
DB 527 GCGATTACCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 586
QY 702 GGGGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 761
DB 587 GGGGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 646
QY 762 CCGTGAACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 821
DB 647 GCGTGTACCAATTAATCACTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 706
QY 822 CTGCTTGTGCTCTCTTCAACGAGTGTCTCTCAACGAGGAGGAGGAGGAGGAGGAGGAGGAG 881
DB 707 CTGCTTGTGATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 766
QY 882 GCGACGCTGACACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 941
DB 767 GCGCTTGTCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 826
QY 942 TAGCGTGCCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1001
DB 827 CAGCGTGCCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 886
QY 1002 GCGCTTGTGAGTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1061
DB 887 CTCCTTTGAGTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 946
QY 1062 GAGTCTCAGTATGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1121
DB 947 GATCTTCACATATCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1006
QY 1122 CT---GAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1178
DB 1007 CTGAGCAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1066
QY 1179 GCGCGCTCTCCAGACTTAATAATGATACCACTTAACCTGTGAGGAGGAGGAGGAGGAGGAGGAG 1238

```

```

Db      1067 GTTGAAGAGTCCAGACTTAAA--TATTACACTCTCTGTGAAAATACTAGTCCGGA 1123
QY      1239 CTCCTTCCCCGCTTGGAGCATGAGCGCGGAAAGCATGCCCCAGGCTG--GGCA 1297
Db      1124 TTCTACCCCTCTTTGGAGCCCATATGACTTAAGACTGTGTAAAGACGCGTCCAGGA 1183
QY      1298 GGAGAGCTCCAGAGAGGCACTGAGCGCTGTGCGCGAGGCTTCGACATCCGAGCA 1357
Db      1184 GCATAGCTTAGTGTGAGAACTGCTTGTCCCTGCGGAGCAGCAAGGCGGCACTCCATGTG 1243
QY      1358 CC-----AGGAAAGTCTCTGGGCGATCTGTAATTAACCTTTTCTTTGT 1408
Db      1244 CCTTGCTTAGAGAAATGATTTCTCAGAGAACTGTAAATACTTTTGTTCCTTT 1303
QY      1409 TTTTAAAAAAA 1421
Db      1304 TTTTCTTCAAAA 1316

```

RESULT 5

```

US-10-187-657-5
; Sequence 5, Application US/10187657
; Publication No. US2003006831A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy K.W.
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Azimzai, Yajida R.
; TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
; FILE REFERENCE: PV-0009 CIP
; CURRENT APPLICATION NUMBER: US/10/187,657
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: PCT/US00/07817
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/139,565
; PRIOR FILING DATE: 1999-06-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 522
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US2003006831A1 3221661R6
; NAME/KEY: unsure
; LOCATION: 455, 480, 483
; OTHER INFORMATION: a, c, g, or other
US-10-187-657-5

```

```

Query Match      31.3%; Score 451.6; DB 14; Length 522;
Best Local Similarity 96.7%; Pred. No. 1,3e-115;
Matches 502; Conservative 0; Mismatches 12; Indels 5; Gaps 4;

```

```

QY      1 AAAGTAACGGCTACAGCACTGAGAAATAGTTCCGCTCCCGGCTTGAAAACTCTGTG 60
Db      9 AAAGTAACGGCTACAGCACTGAGAAATAGTTCCGCTCCCGGCTTGAAAACTCTGTG 68
QY      61 GTACCAACCCAGAGCGCTTGAAGAGCAAGCCCACTCCAGCTTCTTTAAGAGAGATGCA 120
Db      69 GTACCAACCCAGAGCGCTTGAAGAGCAAGCCCACTCCAGCTTCTTTAAGAGAGATGCA 128
QY      121 GGAATCAGACTTACCAAGCCCACTCGGTCCAGCTTTGAAGCAAGAGAGCCCAAGAC 180
Db      129 GGAATCAGACTTACCAAGCCCACTCGGTCCAGCTTTGAAGCAAGAGAGCCCAAGAC 188
QY      181 GCGGCTTCCGCGGTCAGAGAGCCCAAGCTTGTGTGCTTGGCTGCGGCTGCGTGAAGC 240
Db      189 GCGGCTTCCGCGGTCAGAGAGCCCAAGCTTGTGTGCTTGGCTGCGGCTGCGTGAAGC 248
QY      241 ACTGGCGGCGGTGACAGATGACCTGTGGAACGGGCTACTGCTTTTAAACCCCAAGCCC 300
Db      249 ACTGGCGGCGGTGACAGATGACCTGTGGAACGGGCTACTGCTTTTAAACCCCAAGCCC 308

```

```

QY      301 CCGATAGCCGAGAGCTTACAGGCTTCACTGCTCATGCTAATTTAGTGTGCTCTCA 360
Db      309 CGGATAGCCGAGAGTTCAGAGGCTTCACTGCTCATGCTAATTTAGTGTGCTCTCA 367
QY      361 GGAGCAAGCTTCTCTCATCTTGGCGGGATCCGTGGCCACTGCGCCTGTTTGGTTG 420
Db      368 GGAGCAAG--TTCTGTGATCTTGGCGGGATCCGTGGCCACTGCGCCTGTTTGGTTG 425
QY      421 GTGAGAGTCTTCTCAGTCTGTTCATAGGCGCAGAAAATTGCGCTGTGCACTTCAATGCA 480
Db      426 GTGAGAGTCTTCTCAGTCTGTTCATAGGCGCAGAAAATTGCGCTGTGCACTTCAATGCA 484
QY      481 GAATGTTCTGTGGGTACAGTGAACCAACCAATCTTAC 519
Db      485 GAATGTTCTGTGGGTACAGTGAACCAACCAATCTTAC 522

```

RESULT 6

```

US-10-094-749-1195/c
; Sequence 1195, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHITO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTORYUKI
; APPLICANT: NAGAHARA, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 08435/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1195
; LENGTH: 2684
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-1195

```

```

Query Match      27.5%; Score 396; DB 15; Length 2684;
Best Local Similarity 97.6%; Pred. No. 8.8e-100;
Matches 402; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```

```

QY      1007 TCTGGTACAGCTGGGAACCGGCGTCTGTGCTTCTTCTGAGAGGCGGTGATGATC 1066
Db      2363 TTTCGATATCCCAACCGGCAAGGCGTCTGTGCTTCTTCTGAGAGGCGGTGATGATC 2304
QY      1067 TCCAGTATGTTGGGCCCAAGCGCTCTTGCACCCCTTCTGACCAAGCCCAAGAGACTGCA 1126
Db      2303 TCCAGTATGTTGGGCCCAAGCGCTCTTGCACCCCTTCTGACCAAGCCCAAGAGACTGCA 2244
QY      1127 GCCAGAGAGAGGGGGCTACCTCTTATCTGTGGGAGACCACTGACCAAGAGAGCGGCTC 1186
Db      2243 GCCAGAGAGAGGGGGCTACCTCTTATCTGTGGGAGACCACTGACCAAGAGAGCGGCTC 2184

```

QY 1187 TCCGAGCTTAAATGATATACCACTAACCCTGTGAGGGAGGACCAATCTGAGCTCTCC 1246
DB 2183 TCCGAGCTTAAATGATATACCACTAACCCTGTGAGGGAGGACCAATCTGAGCTCTCC 2124
QY 1247 CCGGCTTGGGACATCGAGGCGGGAAGACAGTGTCCCGCCAGGCTGTGGGCGAGAGAGCTC 1306
DB 2123 CCGGCTTGGGACATCGAGGCGGGAAGACAGTGTCCCGCCAGGCTGTGGGCGAGAGAGCTC 2064
QY 1307 CAGGAAGGAGCACTAGGAGCTGTGGCGGAGGCGCTGAGCAATCCGAGGACACAGAGAAA 1366
DB 2063 CAGGAAGGAGCACTAGGAGCTGTGGCGGAGGCGCTGAGCAATCCGAGGACACAGAGAAA 2004
QY 1367 GTCTCTGGGCGATCTGTAAATTAACCTTTTCTTTTGTTTTAAAAA 1418
DB 2003 GTCTCTGGGCGATCTGTAAATTAACCTTTTCTTTTGTTTTAAAAA 1952

RESULT 7

US-10-187-657-7
; Sequence 7, Application US/10187657
; Publication No. US2003006831A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy K.W.
; APPLICANT: Bauch, Mariah R.
; APPLICANT: Azimzai, Yalda
; TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
; FILE REFERENCE: PV-0009 CIP
; CURRENT APPLICATION NUMBER: US/10/187,657
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: PCT/US00/07817
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/119,565
; PRIOR FILING DATE: 1999-06-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program
; SEQ ID NO 7
; LENGTH: 450
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US2003006831A1 SBQA04510D1
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 140, 165, 203, 237, 269, 307, 338, 344, 347, 354, 390, 430, 445
; OTHER INFORMATION: a, t, c, g, or other
US-10-187-657-7

Query Match 26.8%; Score 385.8; DB 14; Length 450;
Best Local Similarity 93.7%; Pred. No. 2,7e-97;
Matches 404; Conservative 0; Mismatches 26; Indels 1; Gaps 1;

QY 616 CTGAACGAGACCATGATCAACAAGAGAGTTCCTGGGCTCGAAAGAGATTACGCC 675
DB 21 CTGAACGAGACCATGATCAACAAGAGAGTTCCTGGGCTCGAAAGAGATTACGCC 80
QY 676 GCGGAGTACGGAACGCACTGGAAGAGGGGCTGCCGAGCCAGTGTCTTACTTGGCGAG 735
DB 81 GCGGAGTACGGAACGCACTGGAAGAGGGGCTGCCGAGCCAGTGTCTTACTTGGCGAG 140
QY 736 AAGTTCACACGAGAGAGCCCTTGGGCTGTACACACAGTACACACCTGGGAGGACACTAC 795
DB 141 AAGTTCACACGAGAGAGCCCTTGGGCTGTACACACAGTACACACCTGGGAGGACACTAC 200
QY 736 GCGTGGGCGACGCTATGAGGTGAGCTTGTCTTGTGCTCTCTCTTCAACGTCGTCTCC 855
DB 201 GCGTGGGCGACGCTATGAGGTGAGCTTGTCTTGTGCTCTCTCTTCAACGTCGTCTCC 260
QY 856 AGCGCGCGCGCTTATGAGAGCGCTGTGACCTGTGACACACCGAGGCTTGGGCTCTTC 915
DB 261 AGCGCGCGCGCTTATGAGAGCGCTGTGACCTGTGACACACCGAGGCTTGGGCTCTTC 320
QY 916 GGGGTCTTGCCTTGAGCTGCATCTTACGCGTGCCTGCGCGCTACGAGCTCC 975

DB 321 GGGGTCTTGCCTTGAGCTGCATCTTACGCGTGCCTGCGCGCTACGAGCTCC 380
QY 976 TCCGAGCTTAAATGATATACCACTAACCCTGTGAGGGAGGACCAATCTGAGCTCTCC 1035
DB 381 TCCGAGCTTAAATGATATACCACTAACCCTGTGAGGGAGGACCAATCTGAGCTCTCC 439
QY 1036 TGCCTTCTCT 1046
DB 440 TGCCTTCTCT 450

RESULT 8

US-09-759-130B-424
; Sequence 424, Application US/09759130B
; Publication No. US20030022279A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: McCarthy, Sean A.
; APPLICANT: Fraser, Christopher C.
; APPLICANT: Sharp, John D.
; APPLICANT: Barnes, Thomas S.
; APPLICANT: Kirst, Susan J.
; APPLICANT: Mackay, Charles R.
; APPLICANT: Myers, Paul S.
; APPLICANT: Leiby, Kevin R.
; APPLICANT: Whitton, Nicolas
; APPLICANT: Goodaarl, Andrew
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
; FILE REFERENCE: MPIO-5350NMIM
; CURRENT APPLICATION NUMBER: US/09/759,130B
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 09/479,249
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/559,497
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/596,194
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/342,364
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/608,452
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/393,996
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: US 09/602,871
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/420,707
; PRIOR FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 450
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 424
; LENGTH: 1029
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-759-130B-424

Query Match 24.6%; Score 354.8; DB 10; Length 1029;
Best Local Similarity 64.5%; Pred. No. 1.8e-88;
Matches 546; Conservative 0; Mismatches 297; Indels 3; Gaps 1;

QY 279 ACTGCTTTTACACCCAGCGCCGAGTGGCGAGGCTTACAGGCTTCACTGCTATGCT 338
DB 21 ATTCCCTTTATGCTGCTGCTCCCAAGCAACTTCCGAGTGAACACACTTTGGCAGAT 80
QY 339 TATTCATGCTTTTGGCTCTTACAGCAAGCTTCTGCTCATCTTGGCGGAGATCCGTGG 398

```

Db      81 CATCATGATCTTTCTGACTGCACTGGCCACGTTCACTGCTGCTGCGATTCGGGG 140
Qy      399 CCACGCGCGCTGTTTGGTTGGTAGAGTTCTTCTCACTGCTGTTCAATAGCCGAGAAAT 458
Db      141 AAAAGAGAGGCTGTTCTGGCTCTTGGGTTGGTAGCAGCTTATTTCACTGGGGCTGCAT 200
Qy      459 TGTGCTGTGCACTTCACTGAGTGAAGATGTTGCTGGGTAAGTGAACACCAACACATCCTA 518
Db      201 CCGGCTGTGAATTTCACTTGTAGTGTCTGTGGCCAGGTCAGACCAACACATCATATA 260
Qy      519 CAAGGCTTCAGGCGCGGCTTACAGCCGCTGTGCTGTGCTGTGGGCTGTGAAGG 578
Db      261 CAAGGCTTCAGTGTGAGTGAATGAGCTGATATGAGCTGTGAGGCTGTGAAGGCTGTGA 320
Qy      579 CATTAAATTTACACTCAAGGAGACCCAGTGCATGAGTGAACGACCATTTGACTCAAA 638
Db      321 AGTCACATCACTCAAGGAGACCCCGTGCAGAGTGAATGACATCAATTTCAAA 380
Qy      639 CGAGCATTTACTGCGCTGTGAAGAGATTTACCGCGGAGTACGGAACGCACTGGA 698
Db      381 CGAGGAGTTCACTGCGCTGTGAGTGAATGAGTGAATGAGTGAAGTGTGCAAGGCTGTGA 440
Qy      699 GAAGGAGCTGCGGAGACCCAGTGCATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 758
Db      441 GAAGGAGCTGCGGAGACCCAGTGTGTGTGCTGCTGAGTGAAGTGAAGTGAAGTGAAGTGA 500
Qy      759 CGGCTGTACCACTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 818
Db      501 TGACCTATACCGCAGTACCGCTGCGGAGACATCACTGAGCATGCTATGAGTGGC 560
Qy      819 GTTCTGCTTGGCTCTCTCCAAAGTGTGCTGCTGCAAGCGCGCGCTGCTGCAAGG 878
Db      561 ATTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 620
Qy      879 CCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 935
Db      621 CTACATGCTATGCGCCACGCGGAGTCTTCCAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 680
Qy      936 CATCTGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 995
Db      681 CACATCACTACCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 740
Qy      996 CGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1055
Db      741 TGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 800
Qy      1056 CGTGTGAGTCTCCAGTATGTTGGCCGAGCGCTTTCGCACTTTCGACCAAGCGC 1115
Db      801 TATGGCGGTGGCCCAAGATGAGCTTCAAGGCTGAAGGCTTTCTTCAACCAAGTGT 860
Qy      1116 CAAGGA 1121
Db      861 GGATGA 866

```

```

RESULT 9
US-10-042-431-54
; Sequence 54, Application US/10042431
; Publication No. US20020182675A1
; GENERAL INFORMATION:

```

```

; APPLICANT: MCCARTHY, Sean A
; APPLICANT: BAENES, Thomas M
; APPLICANT: FRASER, Christopher C
; APPLICANT: SHARP, John D
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING DIAGNOSTIC,
; TITLE OF INVENTION: PREVENTIVE, THERAPEUTIC, AND OTHER USES
; FTE REFERENCE: 10147-602
; CURRENT APPLICATION NUMBER: US/10/042,431
; PRIOR FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24

```

```

; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO: 54
; LENGTH: 1029
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-042-431-54

```

```

Query Match          24.6%; Score 354.8; DB 13; Length 1029;
Best Local Similarity 64.5%; Pred. No. 1.8e-88;
Matches 546; Conservative 0; Mismatches 297; Indels 3; Gaps 1;

```

```

Qy      279 ACTGCTTTTACCCCGAGCCCGCATGCCGAGCTTACGCTTCCATGCTCATCT 338
Db      21 ATTCCCTTCATGCTGAGCCCAAGCCCACTTCCAGTGAACCACTTGGCCAGCAT 80
Qy      339 TATTCAGTGTGTTTGGCTCTAGACGCAAGCTTCTGCTCATCTTTCGCGG3ATCCGTG 398
Db      81 CATCATGATCTTTCTGACTGCACTGCGCCACGCTTACGTATCTGCTGCTGCAATTCGGGG 140
Qy      399 CCATCGCGCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 458
Db      141 AAAAGAGAGGCTGTTCTGGCTCTTGGGTTGGTAGCAGCTTATTTCACTGGGGCTGCAT 200
Qy      459 TGTGCTGTGCACTTCACTGAGTGAAGATGTTGCTGGGTAAGTGAACCAACACATCCTA 518
Db      201 CCGGCTGTGAATTTCACTTGTGAGTGTCTGTGGCCAGGTCAGACCAACACATCATATA 260
Qy      519 CAAAGCTTCAGCGAGCGCGGTTACAGCCGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 578
Db      261 CAAGGCTTCAGTGTGAGTGAATGAGCTGATATGAGTGAAGTGAAGTGAAGTGAAGTGAAG 320
Qy      579 CATTAAATTTACACTCAAGGAGACCCAGTGCATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 638
Db      321 AGTCACATCACTCAAGGAGACCCCGTGCAGAGTGAATGACATCAATTTCAAA 380
Qy      639 CGAGCATTTACTGCGCTGTGAAGAGATTTACCGCGGAGTACGGAACGCACTGGA 698
Db      381 CGAGGAGTTCACTGCGCTGTGAGTGAATGAGTGAATGAGTGAAGTGTGCAAGGCTGTGA 440
Qy      699 GAAGGAGCTGCGGAGACCCAGTGTGTGTGCTGCTGCAAGCGCGCTGCTGCAAGG 758
Db      441 GAAGGAGCTGCGGAGACCCAGTGTGTGTGCTGCTGCAAGCGCGCTGCTGCAAGG 500
Qy      759 CGGCTGTACCACTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 818
Db      501 TGACCTATACCGCAGTACCGCTGCGGAGACATCACTGAGCATGCTATGAGTGGC 560
Qy      819 GTTCTGCTTGGCTCTCTCCAAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 878
Db      561 ATTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 620
Qy      879 CCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 935
Db      621 CTACATGCTATGCGCCACGCGGAGTCTTCCAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 680
Qy      936 CATCTGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 995
Db      741 TGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 800
Qy      1056 CGTGTGAGTCTCCAGTATGTTGGCCGAGCGCTTTCGCACTTTCGACCAAGCGC 1115
Db      801 TATGGCGGTGGCCCAAGATGAGCTTCAAGGCTGAAGGCTTTCTTCAACCAAGTGT 860
Qy      1116 CAAGGA 1121
Db      861 GGATGA 866

```


RESULT 10
US-09-759-130B-423
Sequence 423, Application US/09759130B
Publication No. US20030022279A1
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc.
APPLICANT: McCarthy, Sean A
APPLICANT: Fraser, Christopher C
APPLICANT: Sharp, John D
APPLICANT: Barnes, Thomas S
APPLICANT: Kirt, Susan J
APPLICANT: Mackay, Charles R
APPLICANT: Myers, Paul S
APPLICANT: Leiby, Kevin R
APPLICANT: Wishton, Nicolas
APPLICANT: Goodearl, Andrew
APPLICANT: Holtzman, Douglas A
TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
TITLE OF INVENTION: USES
FILE REFERENCE: MP100-5350NM1M
CURRENT APPLICATION NUMBER: US/09/759,130B
CURRENT FILING DATE: 2002-09-16
PRIOR APPLICATION NUMBER: US 09/479,249
PRIOR FILING DATE: 2000-01-07
PRIOR APPLICATION NUMBER: US 09/559,497
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: US 09/578,063
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: US 09/333,159
PRIOR FILING DATE: 1999-06-14
PRIOR APPLICATION NUMBER: US 09/596,194
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US 09/342,364
PRIOR FILING DATE: 1999-06-29
PRIOR APPLICATION NUMBER: US 09/608,452
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/393,996
PRIOR FILING DATE: 1999-09-10
PRIOR APPLICATION NUMBER: US 09/602,871
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: US 09/420,707
PRIOR FILING DATE: 1999-10-19
NUMBER OF SEQ ID NOS: 460
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 423
LENGTH: 2133
TYPE: DNA
ORGANISM: Homo sapiens
US-09-759-130B-423

Query Match 24.6%; Score 354.8; DB 10; Length 2133;
Best Local Similarity 63.8%; Pred. No. 2.5e-86;
Matches 555; Conservative 0; Mismatches 312; Indels 3; Gaps 1;

QY 255 CAGCATGACCTCTGTGGAAGGCGTACTGCTTTTACCCCGGCGCATGCCGAGG 314
DB 66 CAAGATGGCTACTTTGGACACACATTCCTTTATGCTGCGCCCAAGCAACCTTCCC 125
QY 315 CTTGACGCTTCACGCTCATGCTATTTCTAGTGTGCTTTGAGCTCTAGACAGAGCTTCT 374
DB 126 GATGACACCACTTTGGCCAGCATCATGATCTTTCTGCTGACATGGCCAGCTTCAT 185
QY 375 GCTCATCTTGGCCGAGATCCGTGACCATCGCGCTGTTTGTGTGTAAGTCTTCT 434
DB 186 CGTCACTCCCTGCTGCTATTCGGGGAAGACGAGGCTGTTCTGCTGCTGCGGTGGTAC 245
QY 435 CAGTGTGTTATAGGCGGAGAAATGCTGCTGACCTTCACTGACAGATGTTGCGG 494
DB 246 CAGCTTATTCATCGGGGCTGCAATCTGCTGTGAATTTCACTTCTGAGTGTGTGG 305
QY 495 TACAGTGAACACCAACATCTTACAAAGCTTCAAGCGGAGCGGCTTACAGCCCTGT 554

DB 306 CAGGTGAGCACCACATCATATACAGGCTTCAAGTCTGAGTGAATCAGCGCTGATAT 365
QY 555 CGGCTGCTGCTGGGCTGAGGGGCAATTATTTACACTCAAGGAGACCCAGTGCATCA 614
DB 366 TGGCTGCAAGTGGGCTGGTGAAGTCAACTACACTCAAGGAGACCCCGTGCAGCA 425
QY 615 GCTGACAGACCATTTGACTACAAAGAGAGTTCACTGGCGTCTGAAAAGAAATTCAC 674
DB 426 GCTGAATGAGACCATCAATTACAAAGAGATTCACCTGGCCCTGGGTGAAGATGAC 485
QY 675 CGGAGATGAGGAG 734
DB 486 TGAAGATGTCAGAAAGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 545
QY 735 GAAGTTACACCGAGTAGCCCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 794
DB 546 GAAGTTACTCAAG 605
QY 795 CGCTGAGGAG 854
DB 606 CACCTGAGCAGATGATGAG 665
QY 855 CAGCGCGAG 914
DB 666 CAGCTGCTGCTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 725
QY 915 CG--GGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 971
DB 726 GAGCTGCTGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 785
QY 972 CTCTCGAGCTACACACTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1031
DB 786 CGCTTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 845
QY 1032 CTTGCTCTTCTTCTGAG 1091
DB 846 GCTGT 905
QY 1092 TCGACACCTTGTGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1121
DB 906 GAGGCTTCTTCAACAGAGTGTGATGA 995

RESULT 11
US-10-042-431-53
Sequence 53, Application US/10042431
Publication No. US20020182675A1
GENERAL INFORMATION:
APPLICANT: MCCARTHY, Sean A
APPLICANT: BARNES, Thomas M
APPLICANT: FRASER, Christopher C
APPLICANT: SHARP, John D
TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING DIAGNOSTIC,
TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER USES
FILE REFERENCE: 10147-6U2
CURRENT APPLICATION NUMBER: US/10/042,431
CURRENT FILING DATE: 2001-10-25
PRIOR APPLICATION NUMBER: US 09/333,159
PRIOR FILING DATE: 1999-06-14
PRIOR APPLICATION NUMBER: US 09/578,063
PRIOR FILING DATE: 2000-05-24
NUMBER OF SEQ ID NOS: 79
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 53
LENGTH: 2133
TYPE: DNA
ORGANISM: Homo sapiens
US-10-042-431-53

Query Match 24.6%; Score 354.8; DB 13; Length 2133;
Best Local Similarity 63.8%; Pred. No. 2.5e-86;
Matches 555; Conservative 0; Mismatches 312; Indels 3; Gaps 1;

```

QY 255 CAGCATGACCGCTGGAACGGCGTACTGCTTTTAAACCCCGACCCCGGACATGCGCAG 314
Db 66 CAGATGGCTACTTTGGAGACACATTCCTCTTATGCTGGCCCCCAAGCCACTTCC 125
QY 315 CTTCACGCTTCCACTGCTATCTTATTTAGTGTGTTGGCTTAGCAGCAAGTTCT 374
Db 126 GATGACACCACTTTGGCCAGCATCATGATCTTTGACTGACATGGCCACTTAT 185
QY 375 GCTCATCTTCCCGGAGATCCGTGCACTCGCGCTGGTTTGTGTTGAGATTTCT 434
Db 186 CGTCATCTGCTGGCATTTGGGAAAGAGAGCGCTGTCTGGCTTCGGGTGTAC 245
QY 435 CAGTGTGTATAGGGGCAAAATTTGCTGTGCACTTCAATGCAAGATGTTGTGG 494
Db 246 CAGCTTATTCATGGGGCTGCAATCCTGCTGGAATTTCAATTTGAGTGTGTGG 305
QY 495 TACAGTGAACACCAACATCTCAAAACCTTTAGCGGAGCGGCTTACAGCCGTG 554
Db 306 CAGGTGAGACCAACATCATCAAGAGCTTCACTTGTAGTGGATCAGCGCTGAT 365
QY 555 CGGCTGCTGTGGGCTGAGGGCATTAATTTACTCACAGGAGACCCAGTGCATCA 614
Db 366 TGAGCTGAGGTGGGGCTGGGTGAGTCAACATCACTCAAGGAGACCCCGTGCACA 425
QY 615 GCTGAACGAGACCATGACTACAGAGAGATTCACCTGGCTGAAAGAAATTAAGC 674
Db 426 GCTGAATGAGACCATTAATTAACAAGAGATTCACCTGGCTGGGTGAAATTAAGC 485
QY 675 CGCGAGTACGCGAAGCACTGAGAGAGGGCTGCGGAGCCAGTGTCTTACTTGGCGGA 734
Db 486 TGAGAGTGTGCAAAAGGCTCTGAGAGAGGGCTGCGAGACCTGTGTGTACTGCTGA 545
QY 735 GAAGTTACACCGAGTAGCCCTTGGCGCTGTACACCAAGTACACCTGGCGGAGACTA 794
Db 546 GAAGTTACTTCAAGAGCCCATGTGCTTAAACCGCAGTACCGCTGGCGGAGACTA 605
QY 795 CGGCTGCGGACGCTATGAGTGGCTGTGCTTCTGCTCTCTCAACGTTGCTCTCT 854
Db 606 CACTCAGCATGTATAGGTGGCATTTCTGCTGCTGGCTGCGCAATGTGATCTCTC 665
QY 855 CAGCGCGCCCGCTCTACGAGAGCTTGGCACTGTGACACCGAGCTTGGCGCTTT 914
Db 666 CAGGCTGTGCTGTATATGTTGCTACATGCTATTTGCGCAGGAGATTTTCACTGTT 725
QY 915 CG---GAGCTTGGCGCTTGGCTTCCATCTCTACCGTGGCGCTTGGCGCTAGG 971
Db 726 GGGCTGTGCTTCTTCTTCCATGAGCCATCACTCACTTACCTGTGCTTCCCTTCACTGG 785
QY 972 CTCTCGGCGCTACCACTGATAGCGCGCGCTTCTGAGTCAAGCTGGCAACCGGCT 1031
Db 786 CGCTTCTGTGCTGATCACTCACTATGAGCTTCTGATCACTTACCAAGAGACT 845
QY 1032 CGTGTGCTTCTTCTTGGAGAGGCGCGTGTGATCTCCAGTATGTTGGCGGCGAGGCTCT 1091
Db 846 GCTGTGTGTGTGTGGGCTGTGATTTGGCGTGGCCACAGGATGCACTTCAAGGCT 905
QY 1092 TCGCACCCTTCTGAGCCAAAGCCCAAGGA 1121
Db 906 GAAGGCTTCTTCAACCAAGAGTGTGATGA 935

```

```

; PRIOR APPLICATION NUMBER: PCT/US00/07817
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/139,565
; PRIOR FILING DATE: 1999-06-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program
; SEQ ID NO 6
; LENGTH: 506
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030068311A1 SEQ03652D1
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 87, 89, 93, 95, 124, 178, 227, 253, 262, 279, 285, 287, 293-294, 298,
; LOCATION: 300, 303, 312, 319-321, 323, 325, 327, 332-333, 336, 343-344, 349, 377,
; LOCATION: 405, 409-410, 412, 415, 420, 423, 429, 432, 436-437, 444-445, 448, 452,
; LOCATION: 466, 469-471, 479, 483, 485, 492, 494, 499, 501, 504
; OTHER INFORMATION: a, t, c, g, or other
US-10-187-657-6
Query Match 22.3%; Score 322; DB 14; Length 506;
Best Local Similarity 86.3%; Pred. No. 1.8e-79;
Matches 353; Conservative 0; Mismatches 54; Indels 2; Gaps 2;
QY 464 CTGTGACTTCACTGAGTAATGTTGTGGGTACAGTAAACCAACATCTTCAAAAG 523
Db 17 CTGTGACTTCACTGAGTAATGTTGTGGGTACAGTAAACCAACATCTTCAAAAG 75
QY 524 CCTTACGCGACGCGCGCTTACAGCCCGTGTGCTGCTGCTGAGGAGGCAATTA 583
Db 76 CCTTACGCGACGCGCGCTTACAGCCCGTGTGCTGCTGCTGAGGAGGCAATTA 135
QY 584 ATTTACTCAACAGGAGCCCAAGTGCATCAGCTGAACGAGAACATTAATTAACAGAC 643
Db 136 ATTTACTCAACAGGAGCCCAAGTGCATCAGCTGAACGAGAACATTAATTAACAGAC 195
QY 644 AGTTACCTGGGGGTGTGAAGAGATTTACGCGCGAGTACGCGAGCACTGAGAGAG 703
Db 196 AGTTACCTGGGGGTGTGAAGAGATTTACGCGCGAGTACGCGAGCACTGAGAGAG 255
QY 704 GGTGCGGAGCCAGTGTCTTCACTGCGGAGAAATTTACACGAGTAGCCCTTGGCGCC 763
Db 256 GGTGCGGAGCCAGTGTCTTCACTGCGGAGAAATTTACACGAGTAGCCCTTGGCGCC 315
QY 764 TGTACCAACAGATCAACCTGGGCGGACACTAGGCTCGGCGACGCTATGGGTGGCTTCT 823
Db 316 TGTNNANNCNAGACANNNGGAGCANNATCANCTGGCCACGCTATGGGTGGCTTCT 375
QY 824 GCTTGTGCTCC-TCTCCAAGTGTGCTTCCAGCGCGCGCGCTCT 871
Db 376 GNTTGTGCTCTTCTTCCAGAGGTGTGCTGCTCANNANNGNCGCNCNT 424

```

```

RESULT 12
US-10-187-657-6
; Sequence 6, Application US/10187657
; Publication No. US20030068311A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy K.W.
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Azimzai, Yalda
; TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
; FILE REFERENCE: PV-0009 CIP
; CURRENT APPLICATION NUMBER: US/10/187,657
; CURRENT FILING DATE: 2002-07-01

```

```

RESULT 13
US-10-187-657-3
; Sequence 3, Application US/10187657
; Publication No. US20030068311A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy K.W.
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Azimzai, Yalda
; TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
; FILE REFERENCE: PV-0009 CIP
; CURRENT APPLICATION NUMBER: US/10/187,657
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: PCT/US00/07817
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/139,565
; PRIOR FILING DATE: 1999-06-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program

```

```

; SEQ ID NO 3
; LENGTH: 346
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Inctye ID No. US20030068311A1 1752794F6
; NAME/KEY: unsure
; LOCATION: 91, 188, 206, 291, 337
; OTHER INFORMATION: a, t, c, g, or other
US-10-187-657-3

```

```

Query Match      20.8%; Score 300; DB 14; Length 346;
Best Local Similarity 94.6%; Pred. No. 2.1e-73;
Matches 330; Conservative 0; Mismatches 15; Indels 4; Gaps 2;

```

```

QY 927 CTTGGCCTTCATCTCTAGAGGTGCGCTGCTGCGCGCTGAGCTCTCTCGAGGCTCAC 986
DB 1 CTTGGCCTTCATCTCTAGAGGTGCGCTGCTGCGCGCTGAGCTCTCTCGAGGCTCAC 60
QY 987 CACTAGTACGGGCGCGCTTCTGAGTCAAGCTGCGCAACCGGCGTCTGCTCTTCT 1046
DB 61 CACTAGTACGGGCGCGCTTCTGAGTCAAGCTGCGCAACCGGCGTCTGCTCTTCT 120
QY 1047 CGGAGGGGCGCGTGTGAGTCTTCCAGTATGTTCCGCGCAAGGCTTCTGCGA 1106
DB 121 CGGAGGGGCGCGTGTGAGTCTTCCAGTATGTTCCGCGCAAGGCTTCTGCGA 180
QY 1107 CCAGAGCGCCAGAGACTGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1165
DB 181 CCAGAGCGCCAGAGACTGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
QY 1166 CACTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1225
DB 241 CACTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
QY 1226 GACCCATCTGAGCTCTTCCCGCTTGGAGACATCGAGAGCGGAGAG 1274
DB 301 GACCCATCTGAGCTCTTCCCGCTTGGAGATCGGAGAGCGGAGAG 346

```

RESULT 14

```

US-10-094-749-1195
; Sequence 1195, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAL TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHITO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOMYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14

```

```

; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1195
; LENGTH: 2684
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-1195

```

```

Query Match      17.1%; Score 245.8; DB 15; Length 2684;
Best Local Similarity 67.4%; Pred. No. 7.3e-58;
Matches 362; Conservative 0; Mismatches 172; Indels 3; Gaps 1;

```

```

QY 588 TACACTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 647
DB 250 TGCAATCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 309
QY 648 CACCTGGAGCTGTGAAGAAGATTACGCGCGAGTACGCGAACGCACTGAGAGAGAGAG 707
DB 310 CACCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 369
QY 708 GCGGAGCCCAAGTGTCTTACCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 767
DB 370 GCGGAGCCCAAGTGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAG 429
QY 768 CCACAGATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 827
DB 430 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 489
QY 828 CTGGCTCTCTTCAAGTGTGTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 887
DB 490 CTGGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 549
QY 888 GGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 944
DB 550 ATTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 609
QY 945 GGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1004
DB 610 CACCTCAAGCTGTGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 669
QY 1005 CTTGCGAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1064
DB 670 CTTGCGAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 729
QY 1065 TCTTCAGATGTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1121
DB 730 GAGCCAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 786

```

RESULT 15

```

US-10-027-632-91970
; Sequence 91970, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002

```

Tue Feb 24 07:50:03 2004

us-09-936-456-1.std.rnpb

Page 11

```

P R I O R F I L I N G D A T E : 1999-08-09
N U M B E R O F S E Q I D N O S : 325720
S O F T W A R E : F A S T S E Q f o r W i n d o w s V e r s i o n 4.0.
S R Q I D N O : 91970
L E N G T H : 439
T Y P E : D N A
O R G A N I S M : H u m a n
F E A T U R E :
N A M E / K E Y : m i s c . f e a t u r e
L O C A T I O N : ( 1 ) - - ( 439 )
O T H E R I N F O R M A T I O N : n = A , T , C o r G
U S - 10-027-652-91970

```

Query Match	13.9%	Score 200.6;	DB 15;	Length 439;
Best Local Similarity	97.3%;	Pred. No. 1.3e-45;		
Matches 214; Conservative	0;	Mismatches 5;	Indels 1;	Gaps 1;

QY 812 GGGTGGGGTCTCTCTTGGCTGCTCTTCCAAAGTCTCTCTCTCCACGCGGAGCCCGCTCT 871

Db 133 GGGTGGGGTCTCTCTTGGCTGCTCTTCCAAAGTCTCTCTCTCCACGCGGAGCCCGCTCT 192

QY 872 ACCGAGGCGTGGACATGCTTG-ACCAACGGAGGCGTTTGCGGCTTTCGAGGCTTTCGCTTG 930

Db 193 GCGGAGGCGTGGACNCTGCTGACCAACCGGAGCGCTTTCGAGCTTTCGAGGCTTTCGCTTG 252

QY 931 GCGTTCATCTCTTAGCGTGGCGCTCTTGCCCGCTCTCCGCTTAGGCTCTCTCGGCGTACCACT 990

Db 253 GCGTTCATCTCTTAGCGTGGCGCTCTTGCCCGCTCTCCGCTTAGGCTCTCTCGGCGTACCACT 312

QY 991 CAGTACGAGCGCGCGCTTCTGGGTCACAGCTGGGCAACCGGCG 1030

Db 313 CAGTACGAGGCGCGCGCTTCTGGGTCACAGCTGGGCAACCGGCG 352

```
Search completed: February 23, 2004, 23:36:10
Job time : 564.916 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using ew model

Run on: February 23, 2004, 19:35:23 ; Search time 3666.04 Seconds
(without alignments)
11737.841 Million cell updates/sec

Title: US-09-936-456-1

Perfect score: 1441

Sequence: 1 aaagtaacggcctacagacag.....aaaaaaaaaaaaaaaa 1441

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2753289 seqs, 14931090276 residues 55026578

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estnu:*

5: em_estov:*

6: em_estp1:*

7: em_estro:*

8: em_hic:*

9: gb_est1:*

10: gb_est2:*

11: gb_hic:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pin:*

20: em_gss_vit:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_pig:*

27: em_gss_vit:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	694.6	48.2	1594	AK018569	AK018569 Mus muscu
2	657	45.6	1309	AK008816	AK008816 Mus muscu
3	529.6	36.8	537	AI821606	AI821606 nk08a11.x
4	515	35.7	515	AI826629	AI826629 wk35e04.x

C	5	510	35.4	510	9	AI304327	AI304327 q057h06.x
C	6	489.2	33.9	505	9	AI793184	AI793184 nk08a11.y
C	7	482.4	33.5	486	9	AI660560	AI660560 we68b08.x
C	8	466	32.3	1014	13	BY708714	BY708714 BY708714
C	9	446	31.0	446	9	AI984141	AI984141 wu21c02.x
C	10	445.4	30.9	446	9	AA573825	AA573825 nk08a11.s
C	11	434	30.1	434	9	AI931272	AI931272 wu41h04.x
C	12	414	28.7	414	9	AA593860	AA593860 m191f03.s
C	13	409.4	28.4	411	9	AI983793	AI983793 wu20c09.x
C	14	405	28.1	405	9	AI274929	AI274929 q149c11.x
C	15	400.2	27.8	405	9	AI346155	AI346155 qp34f12.x
C	16	392	27.2	538	13	AI660493	AI660493 we67h02.x
C	17	389.4	27.0	538	13	BX517021	BX517021 BX517021
C	18	385	26.7	573	14	CB854121	CB854121 ut-CF-DU1
C	19	381	26.4	381	9	AI281211	AI281211 qk58e08.x
C	20	375.8	26.1	737	14	CB305399	CB305399 ut-CF-EN1
C	21	373.6	25.9	660	14	CF766127	CF766127 CBS004870
C	22	372.8	25.9	646	13	BY720747	BY720747 BY720747
C	23	367.4	25.5	713	14	CB171450	CB171450 JTD602700
C	24	356.4	24.7	1032	29	AY418362	AY418362 Homo sapi
C	25	349.8	24.3	1026	29	AY418364	AY418364 Mus muscu
C	26	346	24.0	483	9	AI019726	AI019726 ua34c11.x
C	27	335	23.2	348	9	AW050605	AW050605 w219b11.x
C	28	333.4	23.1	690	12	BM977010	BM977010 ut-CF-EN1
C	29	333.4	23.1	896	12	BG968477	BG968477 602835393
C	30	324	22.5	687	14	CB850544	CB850544 ut-CF-EN1
C	31	322.4	22.4	324	12	BM987789	BM987789 ut-H-COO-
C	32	308.8	21.4	871	12	B1526485	B1526485 602925070
C	33	304.8	21.2	919	12	B1412602	B1412602 602990526
C	34	296	20.5	296	9	AI821178	AI821178 nei17a03.y
C	35	295.2	20.5	999	12	B1410133	B1410133 602964073
C	36	290.6	20.2	315	9	AA618335	AA618335 ng15g11.s
C	37	289.4	20.1	895	12	B1413312	B1413312 602986494
C	38	277	19.2	277	9	AI732165	AI732165 nei17a03.x
C	39	275	19.1	608	14	CF108325	CF108325 Shultzom1
C	40	274.4	19.0	654	10	BE914798	BE914798 60167822
C	41	263.2	18.3	766	29	AY418363	AY418363 Pan trogl
C	42	260	18.0	260	9	AW009962	AW009962 ws69h06.x
C	43	257.2	17.8	675	12	B1154672	B1154672 602904653
C	44	252	17.5	252	9	AI262416	AI262416 qk38e04.x
C	45	251.8	17.5	1040	12	B1554670	B1554670 603236624

ALIGNMENTS

RESULT 1

LOCUS AK018569

DEFINITION Mus musculus adult male colon cDNA, RIKEN full-length enriched library, clone:9030623N16 product:hypothetical protein, full insert sequence.

ACCESSION AK018569

VERSION AK018569.1 GI:12658338

KEYWORDS HTC; CAP trapper

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE

1 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)

TITLE JOURNAL MEDLINE PUBMED

2 10349636

REFERENCE
AUTHORS
3 Shibata,K.,Itoh,M., Aizawa,K., Nagaoke,S., Sasaki,N., Carninci,P.,
Kono,H., Akiyama,J., Nishii,K., Kitsumai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Mishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujisake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,S., Watanabe,K.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,Y., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multiplexillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL
MEDLINE
20530913
11078861

PUBMED
REFERENCE
AUTHORS
4 The RIKEN Genome Exploration Research Group Phase II Team and the
PANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

TITLE
JOURNAL
REFERENCE
AUTHORS
5 The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 1594)

JOURNAL
REFERENCE
AUTHORS
Arakawa,T., Aizawa,K., Akahira,S., Akiura,T., Arai,A., Aono,H.,
Adachi,J., Bono,H., Carninci,P., Fukuda,S., Furumishi,K.,
Funuro,M., Hanagaki,T., Hara,A., Hayasun,N., Hiramoto,K.,
Hirooka,T., Horii,F., Imotani,K., Ishii,Y., Itch,M., Iwaw,M.,
Kasubawa,T., Kato,H., Kawai,C., Koijima,Y., Komori,H., Kotoda,M.,
Komura,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishii,K.,
Nomura,K., Nunakazi,R., Ohtno,M., Okazaki,Y., Okido,T., Owa,C.,
Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
Tejima,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T.,
Tsukui,T., Toyota,T., Yamamura,T., Yasunishi,A., Yoshida,K.,
Yoshino,M., Yuramatsum,M. and Hayashizaki,Y.

TITLE
JOURNAL
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tezumi-ku, Yokohama,
Kanagawa 230-0045, Japan [E-mail:genome-res@gsc.riken.go.jp],
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)

COMMENT
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5' GAGGAGAAGAGACATCCAAAGACTCTTTTTCCTTTTTTTTNN 3'], cDNA was
prepared by using trihalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. cDNA went
through one round of normalization to Rot = 10.0 and subtraction to
Rot = 165.2. Second strand cDNA was prepared with the primer
adaptor of sequence [5'
GAGGAGAAGATCGATTGATTAATAATATCATCCCCCCCCCCC 3']. cDNA was cleaved
with BamHI and XhoI. Vector: a modified plasmidscript KS(+) after
bulk excision from lambda FLX I. Cloning sites, 5' end, SalI, 3'
end: BamHI. Host: DH10B

FEATURES
SOURCE
Location/Qualifiers
1..1594
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:9030623N16"
/db_xref="MGI:1894813"
/db_xref="taxon:10090"
/clone="9030623N16"
/sex="male"
/tissue_type="colon"

CDS		/clone_id="MIKEN_full-length enriched mouse cDNA library" 144_...1106 /note="unnamed protein product; hypothetical protein (evidence: ProCrest, decoder, NCBI CDS predictor, longest-ORF) putative" /codon_start=1 /protein_id="BAB1281.1" /db_xref="GI:12858339" /translation="MTAMDVLPFPYQPHASFSVPILLIYLVELSLAASFLLILPG VIGSRWFMFLRVYLISLPIGAELVAHFGSDMVRVTMDMSYKAFSPSRVGVNHHG RCSNRMTLVPRLOILTETIDYNRFPWRMLNRYEYXMALEKIPDVLTLAKR FTSSPGLGHQHLNHAHTAAATLWAFFWIIANALLSMAPRLYGIALITLTAFTLL FGVFAPFSIVSPPCFHRJASSAVLTIFYTPEALWPLATGIYLSLLGGVAVILHTPRS ALRFILDLSVKDSNOAKGNPSVLTINNPQHQLKSPLDLNTLLL"
ORIGIN		
Query Match	48.2%; Score 694.6; Db 11; Length 1594;	
Best Local Similarity	75.7%; Pred. No. 7,3e-85;	
Matches 918;	Conservative 0; Mismatches 279; Indels 16; Gaps 4;	
Oy	222 CTGCCCGCTCGTGTCACACTCGGCGCGGTGCAGCATGACCCTGTGGAACGGGTACT	281
Dd	107 CTCGCTGCTCGGAACAAGTTCTGCAATTCTTGCAAGAATGACTGCTGGAACGGGTCT	166
Oy	282 GCCTTTTACCCCAAGCCCCGAGCATGCGCAGGCTTCAGCGTTCAGTCTCATGCTATAT	341
Dd	167 ACCCTTTTACCCCAAGCCCCAAGCATGCGCAGACTTCAGCGTACCGCTGCTCATGTTAT	226
Oy	342 TCTAGTGTTTTTGGCTCTAGCAGAACGTTCTGTCTCATCTTGC CGGGGATCCGTGGCA	401
Dd	227 CCTGATGTCTTGCTTTGGCTGTACTTCTGTTCATCTTGCCTGGAATCCGTGGCA	286
Oy	402 CTCGCGCTGTTTTGGTGTGAGAGATTCTTCTCACTGTCTCATAGCGCAGAAATGT	461
Dd	287 CTCGCGCTGTTTTGGTGTGAGAGATTCCTTCACTGTCTCATAGCGCAGAAATGT	346
Oy	462 GGCTGTGCACTTCAGTGCAGAAATGTTCTGTGGTAAAGTGAACACCAACACTCTTAA	521
Dd	347 GGCTGTGCACTTCAGTGAAGATGAGTCTGTGGGGAAGTGAACCAACACTCTTAA	406
Oy	522 AGCTTTACGCGAGCGCGCTTACAGCCGTGTGGTCTGCTGATGCGGCTGGAAGGAT	581
Dd	407 AGCTTTACGCTCATCCGCTGTTCAGTGCATGTGGTCTGCAAGTGGGCTGCGGCGT	466
Oy	582 TAATATTACCTCACAGGGAACCCAGTGCATAGCTGAACAGAACATTGACTATAAGA	641
Dd	467 TAACATTACCTCCAGGAACACCGAGCGAGCGCTGAACAGAACATTGACTATAAGA	526
Oy	642 GAAGTTCACTTGCGCTGTGAAGAAGATTACGCGCGAGTAGTCGCAACGCACTGGANA	701
Dd	527 GGCTTTCACTTGGCGCTGGAAGCAATCACCAAGAGTAGTATCAAGCTTGGANA	586
Oy	702 GGGGCTGCGGACCCAGTGTCTACCTGGGGAAGATTCAACCGAAGTAGCCCTTGGG	761
Dd	587 GGAATGCGGACCCAGTGTATATCTGGGGAAGATTCAACCGAAGTAGCTTGGG	646
Oy	762 CCTGTACCAAGTACCACTGCGGGAACCTAGCGCTCGGCCACGTTATGGTGGCGTT	821
Dd	647 GCTGTACCAAGTATCACTGGCGGCTACTATGCGCGAGCAACTGTGGTGGCAAT	706
Oy	822 CTGCTTCTGGCTCTCTTCCAAGTGTGTCTCAAGCGCGCGCCGCTGTAGAGAGCT	881
Dd	707 CTGCTTCTGGATCAATCGCAATGCGTGTCTCCATAGCCGCGCCCACTTAGAGAGCT	766
Oy	882 GGCATGTCGACCAACCGAGCTTCCGCTCTTCCGAGTCTTTCGCTTGGCTCATATTC	941
Dd	767 GGCATTGTCAACAACGATGCTTCAAGCTTCCGAGTCTTTCGCTCATATTC	826
Oy	942 TAGCGTCCGCTTCCGCGCTCGCGCTAAGTCTCTCCGCGCTACACACTAATTAAGGCG	1001
Dd	827 CAGCGTCCGCTTCCGCTTCCGCGCTCGCGCTAAGTCTCTCCGCTCATATTAAGGCG	886

```

QY 1002 CGCCTTCTGAGTCAAGCGGCGTCTCTGATGCTCTTCTCGAGAGGCGCGTGT 1061
DB 887 CTCCTTTGGCTACGCTGGCCACCGGCACTCTGAGCCTCTCTCGAGAGGCGCGTGT 946
QY 1062 GAGCTTCAGATATGTTGGCCCGACGCGCTTTGGCAACCTTTGGACCAAGGCCCAAGA 1121
DB 947 GATTCTCACTATACGCGCCGACGCGCTCTGCTCTTTCTGATCTCAAGTCTCAAGA 1006
QY 1122 CT---GGAGCGAGAGAGAGGGGGCTCACCTCTTATCTCGGCGACCACTGCAACAAGA 1178
DB 1007 CTGTAGCAACAGAGCTTAAGAGAACTCTCTCTACCTCTCAACAACCCGACAGCAAGA 1066
QY 1179 GCGCGCTCTCCAGACTTAATATATCAACAATCCTGTAGAGGGGAGCCCAATCTGGA 1238
DB 1067 GTTGAAGAGTCCAGACTTAA---TATTACCACTCTCTGTAAATTAATCTGATCCGGA 1123
QY 1239 CTCTTCCCGCGCTTTGGGACATGCGAGCGCGGAGAGCACTGCGCCGCGCTG-GGCGA 1297
DB 1124 TTCTACCCCTCTTTGGGACCCATGACCTGAAGACTGTGTAAAGAGCGCTGCGAGGA 1183
QY 1298 GGAAGCTCCAGAGAGGCGACTGAGCGCTGCTGGCGCGAGCGCTCGACATCCGAGGCA 1357
DB 1184 GCATAGCTTAGCTGAGAGACTCTCTGTCCCTGCGGAGACGACAGAGGGGCGACCCATGTG 1243
QY 1358 CC-----AGGAAAGTCTCTTGGGCGCATCTGTAAATTAACCTTTTCTTTTGT 1408
DB 1244 CTTTGGCTTAGAGAAATGATTTCTCAGAGAACTGTAAATTAACCTTTTCTTTT 1303
QY 1409 TTTTAAAAAAA 1421
DB 1304 TTTTCTTCAAA 1316

RESULT 2
AK008816 1309 bp mRNA linear HTC 20-SEP-2003
LOCUS AK008816
DEFINITION Mus musculus adult male stomach cDNA, RIKEN full-length enriched
library, clone:2210403N03 product:hypothetical protein, full insert
sequence.
ACCESSION AK008816
VERSION AK008816.1 GI:12843238
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
JOURNAL 99279253
MEDLINE 10349636
PUBMED 10349636
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL 20499374
MEDLINE 11042159
PUBMED 11042159
AUTHORS 3
Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,
Kono, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M.,
Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL 20530913
MEDLINE 1757-1771 (2000)
AUTHORS 10 (11), 1757-1771 (2000)

```

```

PUBMED 11076861
REFERENCE
4 The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
JOURNAL 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 1309)
AUTHORS Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Kono, H.,
Aizawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,
Furumori, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,
Hirooka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,
Kasukawa, T., Kato, H., Kawat, T., Kojima, Y., Kono, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K.,
Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C.,
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y.,
Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T.,
Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K.,
Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
TITLE Direct Submission
JOURNAL Submitted (10-UTL-2000) Yoshinori Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
CDS
FEATURES
source
1..1309
location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:2210403N03"
/db_xref="MG1:1893739"
/db_xref="taxon:10090"
/clone="2210403N03"
/sex="male"
/tissue="stomach"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
142..1104
/note="unannotated protein product: hypothetical protein
(evidence: Procest, decoder, NCBI CDS
Predictor, longest-ORF)
putative"
/codon_start=1
/protein_id="BAB25910.1"
/db_xref="GI:12843239"
/translation="MTAMGCVTPYPPORHARSFVPLIVLPLSLAASFPLIPG
IRHSHRFLVRLSLPLFGAIVAVHSGDMFVBEWNTNSYKAFSSRPQVGVGLK
VGIAGVNTLRGTPQQLNEDIVKERTFLNEDYETVEYHAERKIPDVLIAEK
FPSSCGIHQHYHLPGHYAAATVWVGILLDHRCCALHARPLRGRGFHNRCLRA
LCIRFLRDLFORAALPLPWRIRLHALRRRLILLTLAGIILLGAVVILHYTRPS

```

ORIGIN ALRSLDLVMDCSNOKANSPLTINNPHQELKSPDLNTTLL"

Query Match 45.6%; Score 657; DB 11; Length 1309;
Best Local Similarity 74.5%; Pred. No. 9.9e-80;
Matches 898; Conservative 0; Mismatches 290; Indels 17; Gaps 5;

222 CTGCCCCGCTGGCTGACACCTCGCCGCGCTGACGATGACCTGTGAAACCGGCTACT 281
Db CTGCGCTGCTGGAAACATTTCTGAACTTTCTGAGATGACCTGCGGTGGAGCGGGTGTCT 164
QY 282 GCTTTTAAACCCCGCCGCGCATGCGGAGGCTTCAAGGTTTCACTGCTCATCTGTAT 341
Db 165 ACCCTTTTAAACCCCGCCGCGCATGCGGAGGCTTCAAGGTTTCACTGCTCATCTGTAT 224
QY 342 TCTAGTGTGTTTGGCTTACGAGCAAGCTTCTGCTCATCTTGGCCGGAGATCCGTGCGCA 401
Db 225 CTTGATGTTCTTGTCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 284
QY 402 CTGCGCGCTGGTTTGGTTGGTGAAGTTCTTCTGAGTCTGTTCTAGGCGCGCAAAATGT 461
Db 285 CTGCGCGCTGGTTTGGTTGGTGAAGTTCTTCTGAGTCTGTTCTAGTCTGCGCAAAATGT 344
QY 462 GCTGTGCACTTCACTGAGAGATGCTTGTGAGAGAGATGAGCAACATCTTACAA 521
Db 345 GCTGTGCACTTCACTGAGAGATGCTTGTGAGAGAGATGAGCAACATCTTACAA 404
QY 522 AGCCTTCAAGCGAGCGCGCTTACAGCCGCTGTGCTGCTGCTGCGCTGAGAGGCGAT 581
Db 405 AGCCTTCAAGCGAGCGCGCTTACAGCCGCTGTGCTGCTGCTGCGCTGAGAGGCGAT 464
QY 582 TAAATATTAACATCAAG 641
Db 465 TAAATATTAACATCAAG 524
QY 642 GAGATTCACCTGCGCTTGAAGAGATTAAGCGCGAGTACCGAGCAACGCACTGAGAA 701
Db 525 GAGATTCACCTGCGCTTGAAGAGATTAAGCGCGAGTACCGAGCAACGCACTGAGAA 584
QY 702 GGGGCTGCGCGAG 761
Db 585 GGGGCTGCGCGAG 644
QY 762 CCTGTACACAGATACACCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 820
Db 645 GCTGTACACAGATACACCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 704
QY 821 TGTGCTTGTGCTCTCTTCAAGAGTGTGCTTCAAGCGCGCGCTCTTCAAGAGAGAG 880
Db 705 TGTGCTTGTGCTCTCTTCAAGAGTGTGCTTCAAGCGCGCGCTCTTCAAGAGAGAG 764
QY 881 TGGCACTGCTGACACCGAGAGCTTGGCGCTTGGCGAGTCTTGGCGCTTGGCGCTTGG 940
Db 765 TGGTGTGCTGACACCGAGAGTGTGCTTGGCGCTTGGCGAGTCTTGGCGCTTGGCG 824
QY 941 CTAGCGTGCCTGCTGCGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCT 1000
Db 825 CCAAGCTGCGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCT 884
QY 1001 CCGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCT 1060
Db 885 CCGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCT 943
QY 1061 TGAATCTCAAGTATGTTGCGCGCGAGCGCTTCTTGGCACTTCTTGGCAACCGCAAG 1120
Db 944 TGAATCTCAAGTATGTTGCGCGCGAGCGCTTCTTGGCACTTCTTGGCAACCGCAAG 1003
QY 1121 ACTTCTCAAGTATGTTGCGCGCGAGCGCTTCTTGGCACTTCTTGGCAACCGCAAG 1177
Db 1004 ACTTCTCAAGTATGTTGCGCGCGAGCGCTTCTTGGCACTTCTTGGCAACCGCAAG 1063
QY 1178 AGGCGCTCTCCAGACTTAAATGATGATCACTTAACCTGTGAGGGGAGACCAATCTGG 1237

Db 1064 AGTTGAAGAGTCCAGACTTAAATTAACCACTCTCTCTGTAATAAATACTGACTCGG 1120
QY 1238 ACTCTTCCCGGCTTGGAGCATTCGAGGCGCGGAGAGAGTCCCGCAAGGCTGGGCA 1297
Db 1121 ATTCTTACCCCTCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1180
QY 1298 GAGAGAGTCCAG 1357
Db 1181 AGCATAGCTTAGCTGAG 1240
QY 1358 CC-----AGGAAAGCTCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1408
Db 1241 CTTTGTCTTGAAGATGATTTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1300
QY 1409 TTTT 1413
Db 1301 TTTT 1305

RESULT 3
Al821606/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 537)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Other ESTs: nk08a1.y5
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Stratagene, Inc., David B. Kitzman,
Ph.D.
CDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www.bio.lnl.gov/bdip/image/image.html

FEATURES
source
1..537
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1012892"
/tissue_type="tumor"
/lab_host="SOLR (kanamycin resistant)"
/clone_id="NCI CGAP Co2"
/note="Organ: colon; Vector: Bluescript SK-; Site 1:
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Bulk colon villous adenoma. 5' adaptor sequence:
5' GAATTCGACGAG 3' 3' adaptor sequence: 5'
CTCAGTTTCTTTTCTTTTCTTTT 3' Average insert size: 1.1 kb."

ORIGIN

Query Match 36.8%; Score 529.6; DB 9; Length 537;
 Best Local Similarity 99.3%; Pred. No. 2.8e-62;
 Matches 532; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 878 GCTTGACATGCTGACACCGAGACCTTGGCCCTTTGGGGGTCTTGGCCTTGGCCTTCA 937
 DB 537 GCTTGACATGCTGACACCGAGACCTTGGCCCTTTGGGGGTCTTGGCCTTGGCCTTCA 478
 QY 938 TCTTAGCGTGCCTCTGCTCCGCTCCGCTAGGCTCTCCGCGCTCACTGACAGTACG 997
 DB 477 TCTTAGCGTGCCTCTGCTCCGCTCCGCTAGGCTCTCCGCGCTCACTGACAGTACG 418
 QY 998 GCGCGCTTCTGGGTACGCTGCGACACCGCGCTCTGCTCTTCTCTGAGAGGCGCG 1057
 DB 417 GCGCGCTTCTGGGTACGCTGCGACACCGCGCTCTGCTCTTCTCTGAGAGGCGCG 358
 QY 1058 TGGTGAAGTCCAGTATGTTTCCGGCCAGGCTCTTCCGACCTTCTGACCAAGCGCA 1117
 DB 357 TGGTGAAGTCCAGTATGTTTCCGGCCAGGCTCTTCCGACCTTCTGACCAAGCGCA 298
 QY 1118 AGGACTGACACCGAGAGAGGCGGCTCACTCTTATCCTCGGACACCACTGACAAAGC 1177
 DB 297 AGGACTGACACCGAGAGAGGCGGCTCACTCTTATCCTCGGACACCACTGACAAAGC 238
 QY 1178 AGGCGCTCTCCAGACTTAAATGATATACCACTTACCTGAGAGGAGACCAATCTGG 1237
 DB 237 AGGCGCTCTCCAGACTTAAATGATATACCACTTACCTGAGAGGAGACCAATCTGG 178
 QY 1238 ACTCTTCCCGCTTGGGACATCGACAGGCGCGGAGACAGTCCCGCGACGCTGGGCGCA 1297
 DB 177 ACTCTTCCCGCTTGGGACATCGACAGGCGCGGAGACAGTCCCGCGACGCTGGGCGCA 118
 QY 1298 GGAAGCTCCAGAGAGGCACTGAGCGCTGCTGGCGGAGGCTTGGACATCCGAGGCA 1357
 DB 117 GGAAGCTCCAGAGAGGCACTGAGCGCTGCTGGCGGAGGCTTGGACATCCGAGGCA 58
 QY 1358 CCAAGGAAAGTCCCGGGCGCATCTGTAATTAACCTTTTCTTTTGTATT 1413
 DB 57 CCAAGGAAAGTCCCGGGCGCATCTGTAATTAACCTTTTCTTTTGTATT 2

RESULT 4
 A1826629/c 515 bp mRNA linear EST 21-DEC-1999
 LOCUS wk35604.x1 NCI_CGAP_P122 Homo sapiens cDNA clone IMAGE:2417406 3',
 DEFINITION mRNA sequence.
 ACCESSION A1826629
 VERSION A1826629
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 515)
 NCBI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CCAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/BLN at:
www.bio.linn.gov/bdrip/image/image.html
 Insert Length: 910 Std Error: 0.00
 Seq primer: -40UP from GIBCO
 High quality sequence stop: 467.
 Location/Qualifiers

source 1..515
 /organism="Homo sapiens"
 /mol_type="RNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2417406"
 /sex="male"
 /tissue_type="normal prostate"
 /lab_host="DH10B"
 /clone_id="NCI_CGAP_P122"
 /note="Organ: prostate; Vector: pT713D-Pac (Pharmacia)
 with a modified polylinker; 1st strand cDNA was prepared
 from normal prostate bulk tissue, and was then primed with
 a Not I - oligo(dT) primer. Double-stranded cDNA was
 ligated to Eco RI adaptors (Pharmacia), digested with Not
 I and cloned into the Not I and Eco RI sites of the
 modified pT713 vector. Library is normalized, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

Query Match 35.7%; Score 515; DB 9; Length 515;
 Best Local Similarity 100.0%; Pred. No. 2.7e-60;
 Matches 515; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 902 CCTTGCGGCTCTTGGGAGTCTGCGCTTGGCTTCATCTAGCGTCCGCTGCGCGC 961
 DB 515 CTTTGCGGCTCTTGGGAGTCTTGGCTTGGCTTCATCTAGCGTCCGCTGCGCGC 456
 QY 962 TCGGCTAGGCTCTTGGGAGTCTGCGCTTGGCTTCATCTAGCGTCCGCTGCGCGC 1021
 DB 455 TCGGCTAGGCTCTTGGGAGTCTGCGCTTGGCTTCATCTAGCGTCCGCTGCGCGC 396
 QY 1022 CAACGGGAGTCTGAGGCTCTTGGGAGTCTGCGCTTGGCTTCATCTAGCGTCCGCTGCGCGC 1081
 DB 395 CAACGGGAGTCTGAGGCTCTTGGGAGTCTGCGCTTGGCTTCATCTAGCGTCCGCTGCGCGC 336
 QY 1082 CAGGCGCTTGGGAGTCTGAGGCTCTTGGGAGTCTGCGCTTGGCTTCATCTAGCGTCCGCTGCGCGC 1141
 DB 335 CAGGCGCTTGGGAGTCTGAGGCTCTTGGGAGTCTGCGCTTGGCTTCATCTAGCGTCCGCTGCGCGC 276
 QY 1142 GCTCACTTATCTTGGGAGTCTGAGGCTCTTGGGAGTCTGCGCTTGGCTTCATCTAGCGTCCGCTGCGCGC 1201
 DB 275 GCTCACTTATCTTGGGAGTCTGAGGCTCTTGGGAGTCTGCGCTTGGCTTCATCTAGCGTCCGCTGCGCGC 216
 QY 1202 GTATCACCACTTATCTTGGGAGTCTGAGGCTCTTGGGAGTCTGCGCTTGGCTTCATCTAGCGTCCGCTGCGCGC 1261
 DB 215 GTATCACCACTTATCTTGGGAGTCTGAGGCTCTTGGGAGTCTGCGCTTGGCTTCATCTAGCGTCCGCTGCGCGC 156
 QY 1262 GCAAGCGGAGAGAGTGGCGCGCTGAGGCTCTTGGGAGTCTGCGCTTGGCTTCATCTAGCGTCCGCTGCGCGC 1321
 DB 155 GCAAGCGGAGAGAGTGGCGCGCTGAGGCTCTTGGGAGTCTGCGCTTGGCTTCATCTAGCGTCCGCTGCGCGC 96
 QY 1322 GCGCTGCTGGCGGAGGCTCTGAGGCTCTTGGGAGTCTGCGCTTGGCTTCATCTAGCGTCCGCTGCGCGC 1381
 DB 95 GCGCTGCTGGCGGAGGCTCTGAGGCTCTTGGGAGTCTGCGCTTGGCTTCATCTAGCGTCCGCTGCGCGC 36

RESULT 5
 A1304327/c 510 bp mRNA linear EST 01-FEB-1999
 LOCUS A1304327
 DEFINITION q057106.x1 NCI CGAP C08 Homo sapiens cDNA clone IMAGE:1912667 3',
 similar to contains TAR1.3 TAR1 repetitive element ;, mRNA
 sequence.
 ACCESSION A1304327
 VERSION A1304327
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 510)
 NC-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 TITLE Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmer-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NC-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/ILNt at:
www.bio.lnlnl.gov/bbrp/image/image.html
 Insert Length: 1487 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 476.

FEATURES

source

1..510
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1912667"
 /tissue_type="adenocarcinoma"
 /lab_host="PDH10B"
 /clone_lib="NCI CGAP Co2"
 /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 colon adenocarcinoma, and was then primed with a Not I -
 oligo(dT) primer. Double-stranded cDNA was ligated to Eco
 RI adaptors (Pharmacia), digested with Not I and cloned
 into the Not I and Eco RI sites of the modified pT733
 vector. Library is normalized. Library was constructed by
 Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 35.4%; Score 510; DB 9; Length 510;
 Best Local Similarity 100.0%; Pred. No. 1,3e-59;
 Matches 510; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 904 TTGCGCTCTTGGGGGTTTGGCTTGGCTTCCATCTCTAGACGTCGCTGCGCGCTC 963
 DB 510 TTGCGCTCTTGGGGGTTTGGCTTGGCTTCCATCTCTAGACGTCGCTGCGCGCTC 451
 QY 964 GCGCTAGGCTCTCGCGCTCAGCACTAGTACGGCGCGCTTCTGGGTACGCTGGCA 1023
 DB 450 GCGCTAGGCTCTCGCGCTCAGCACTAGTACGGCGCGCTTCTGGGTACGCTGGCA 391
 QY 1024 ACCGGCGTCTGCTCTTCTCTCGAGAGGGCGTGTGATCTCCAGTATGTTCCGCC 1083
 DB 390 ACCGGCGTCTGCTCTTCTCTCGAGAGGGCGTGTGATCTCCAGTATGTTCCGCC 331
 QY 1084 ACCGGCTTTCGACCTTCTGACCAAGCGCCCAAGACGCAAGCGAGAGAGGGGGG 1143
 DB 330 ACCGGCTTTCGACCTTCTGACCAAGCGCCCAAGACGCAAGCGAGAGAGGGGGG 271
 QY 1144 TCACCTCTTATCTCTCGAGCACTGCAAGAGAGCGCTCTCCAGACTTAAATGT 1203
 DB 270 TCACCTCTTATCTCTCGAGCACTGCAAGAGAGCGCTCTCCAGACTTAAATGT 211
 QY 1204 ATCACCACCTAATCTGTAGAGGGGACCCTATCTGACTCTTCCCGGCTTGGACATCGC 1263
 DB 210 ATCACCACCTAATCTGTAGAGGGGACCCTATCTGACTCTTCCCGGCTTGGACATCGC 151
 QY 1264 AGCGCGGAGAGAGTGGCGCGGCTGAGGCGAGAGAGCTCCAGGAAGGGGCACTAGC 1323
 DB 150 AGCGCGGAGAGAGTGGCGCGGCTGAGGCGAGAGAGCTCCAGGAAGGGGCACTAGC 91
 QY 1324 GCTGCTGGCGAGAGGCTCTGACATCTCGGAGGCACTAGGAAAGTCTTCTGGGGCATCT 1383
 DB 90 GCTGCTGGCGAGAGGCTCTGACATCTCGGAGGCACTAGGAAAGTCTTCTGGGGCATCT 31

QY 1384 GTAATTAACCTTTTCTTTTGTGTTT 1413
 DB 30 GTAATTAACCTTTTCTTTTGTGTTT 1

RESULT 6
 LOCUS AT1791844
 DEFINITION nk08a11.y5 NCI CGAP Co2 Homo sapiens CDNA clone IMAGE:1012892 5'
 similar to contains Alu repetitive element; mRNA sequence.
 ACCESSION AT1791844
 VERSION AT1791844.1 GI:5339486
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 505)
 NC-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 TITLE Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Other ESTs: nk08a11.s1
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
 Emmer-Buck, M.D., Ph.D.
 CDNA Library Preparation: Stratagene, Inc., David B. Krizman,
 Ph.D.

CDNA Library Arraying: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NC-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/ILNt at:
www.bio.lnlnl.gov/bbrp/image/image.html

This read is a RESEQUENCE of a previously sequenced human clone
 Original clone citation: see original entry for original citation
 information
 This 5' resequenced clone has no previous 5' data to verify this
 new read against
 Putative full length read
 The vector to vector length is 671
 Insert Length: 741 Std Error: 0.00
 Seq primer: -40RP from Gibco
 High quality sequence stop: 418.

FEATURES

source

1..505
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1012892"
 /tissue_type="tumor"
 /lab_host="SOLR (kanamycin resistant)"
 /clone_lib="NCI CGAP Co2"
 /note="Organ: colon; Vector: Bluescript SK-; Site 1:
 EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
 Oligo dT. Bulk colon villous adenoma. 5' adaptor sequence:
 5' GAATTCGACGAG 3' 3' adaptor sequence: 5'
 CTCGAGTTTCTTTTCTTTTCTTTT 3' Average insert size: 1.1 kb."

ORIGIN

Query Match 33.9%; Score 489.2; DB 9; Length 505;
 Best Local Similarity 99.2%; Pred. No. 8.3e-57;
 Matches 502; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 736 AAGTTCACCCGATAGCCCTTGGGCTTATACACAGTACCACTGGCGGACACTAC 795
 DB 1 AAGTTCACCCGATAGCCCTTGGGCTTATACACAGTACCACTGGCGGACACTAC 60
 QY 796 GCTTCGACAGCTATAGGTGGGCTTGTGCTCTTGGCTCTCTCCAGAGTGGCTCTCC 855
 DB 61 GCTTCGACAGCTATAGGTGGGCTTGTGCTCTTGGCTCTCTCCAGAGTGGCTCTCC 120

QY	856	ACGCGGAGCCCGCCTCTACGGAGAGCCTGGACAATGTGTACACGCGAGCCTTCGCGCTTC	915
Db	121	AGCGCGGAGCCCGCCTCTACGGAGAGCCTGGACAATGTGTACACGCGAGCCTTCGCGCTTC	180
QY	916	GGGGNTTCGCTTGAGCTTGACCTCACTCAAGCGTGCCTCTGCCCCGTCTCCGCTAGAGTCC	975
Db	181	GGGGTCTTGAGCCTTGAGCCTCACTCTCAAGCGTGCCTCTGCCCCGTCTCCGCTAGAGTCC	240
QY	976	TCCGCGCTCAACCATCAGTACGAGCGCGCGCTTCTGGGTACAGTGGCAACCGGCTCTCG	1035
Db	241	TCCGCGCTCAACCATCAGTACGAGCGCGCGCTTCTGGGTACAGTGGCAACCGGCTCTCG	300
QY	1036	TGCGTCTTCCTCGAGAGGAGCGGTGATGTCCAGATATGTTCGGCCCGACGCGCTTC	1095
Db	301	TGCGTCTTCCTCGAGAGGAGCGGTGATGTCCAGATATGTTCGGCCCGACGCGCTTC	360
QY	1096	ACCCTTCTGAGCAAAAGCGCCAAAGATCTGCAGCGAGAGAGAGGAGGCTCACTCTTATC	1155
Db	361	ACCCTTCTGAGCAAAAGCGCCAAAGATCTGCAGCGAGAGAGAGGAGGCTCACTCTTATC	420
QY	1156	CTGGGAGACCCACTGACAAAGAGGCGCGCTTCTCCAGACTTAAATGATACCACTAAC	1215
Db	421	CTGGGAGACCCACTGACAAAGAGGCGCGCTTCTCCAGACTTAAATGATACCACTAAC	479
QY	1216	CTGTGAGGGAGACCCCATCTGGATCC	1241
Db	480	CTGTGAGGGAGACCCCATCTGGATCC	505

RESULT 7	LOCUS	DEFINITION
AI660560/c	AI660560	486 bp mRNA linear EST 18-DEC-1999
	w686808.x1	Soares Dieckgrafe colon NHDD Homo sapiens cDNA clone
	IMAGE:2346531.3	similar to contains TARI.t3 TARI repetitive element ; , mRNA sequence.

ACCESSION	AI660560	GI:4764130
VERSION	AI660560.1	
KEYWORDS	EST.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 486)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
COMMENT

JOURNAL
COMMENT
Unpublished (1997),
Contact: Robert Strausberg, Ph.D.
Email: csaapbs-r@mail.nih.gov

This clone is available royalty-free through LINTL ; contact the
IMAGE Consortium (info@image.lintl.gov) for further information.
Insert March 1433 03-03-2000

Insert Length: 1433 Std Error:
Seq primer: -40UP from Gibco

FEATURES

- High quality sequence stop: 447.
- Location/Qualifiers

```

high quality seqop: 44.1
FEATURES      Location/Qualifiers
source        1. .486

```

```
source
1. .486
/organism="Homo sapiens
/mol type="mRNA"
```

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/cd_name="TMCE-3245231"

```

```

/clone="IMAGE:2346231"
/tissue_type="colonic m

```

```

disease="
/lab_host="DH10B (phage

```

```

/ iad_host="DHIOB (phage
/ clone_lib="Soares_Diec
/ note="Organ: colon; Ve

```

```
/note="Organ: colon; Ve
modified polylinker; si
strand cDNA was primed
```

strand cDNA was primed
TGTACCAATCTGAAGTGGAGCC

double-stranded cDNA was (Pharmacia), digested with

and Eco RI sites of the
went through one round

went through one round provided by Dr. Brian D

ORIGIN

```

Query Match      33.5%; Score 482.4; DB 9; Length 486;
Best Local Similarity 99.4%; Pred. No. 7e-56;
Matches 483; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

dieck@im.wustl.edu); colonic mucosa represents a range of disease involvement from moderate to severe Crohn's disease; samples include both perforating (fistulas) and non-perforating samples. Library constructed by Bento Soares and M. Fatima Ronaldo. "

QY	943	AGCGTGCAGCTTGC	CCGCTCCG	CTCCGCTTGC	AGTGCCTCTCCG	CGCTACACATCAGTACGG	CGCC	1102
Db	486	AGGCTACGCGCTTGC	CCGCTCCG	CTCCGCTTGC	AGTGCCTCTCCG	CGCTACACATCAGTACGG	CGCC	427
QY	1003	GCCCTTCTGGGTCAC	GCTTGG	CAGCCG	CGCGCTG	TGCTCTTCTCTCGAAGGGCCGTGTG	1062	
Db	426	GCTTTTGGGTCAC	GCTTGG	CAGCCG	CGCGCTG	TGCTCTTCTCTCGAAGGGCCGTGTG	367	
QY	1063	AGCTCCAGTATGTT	TCGGCC	CCAGGCGCTTC	CGACCTCTTG	AGACCAAAAGGCCCAAGAC	1122	
Db	366	AGCTCCAGTATGTT	TCGGCC	CCAGGCGCTTC	CGACCTCTTG	AGACCAAAAGGCCCAAGAC	307	
QY	1123	TGCAGCAGAGAGAG	GGGGCTCA	CTCTTATCCT	CGGACCCA	CTGCAAGACGACC	1182	
Db	306	TGCAGCAGAGAGAG	GGGGCTCA	CTCTTATCCT	CGGACCCA	CTGCAAGACGACC	247	
QY	1183	GCTCTCCCAAGATT	AAAA	TGATATAC	CACTAACCTGT	AGGGGGAGCCCAATGTGACCTCC	1242	
Db	246	GCTCTCTCCCAAGATT	AAAA	TGATATAC	CACTAACCTGT	AGGGGGAGCCCAATGTGACCTCC	187	
QY	1243	TTTCCCGCCTTG	AGACATCG	CAGGCCCG	AGAGCAGTGC	CGCGCAGGCTGGGCGCAGAGAG	1302	
Db	186	TTTCCCGCCTTG	AGACATCG	CAGGCCCG	AGAGCAGTGC	CGCGCAGGCTGGGCGCAGAGAG	127	
QY	1303	GCTTCAGAGAGG	GCATCTAG	CGCTGCTGC	CGCGCATCTG	GAACATCCGACAGGCACCAAG	1362	
Db	126	GCTTCAGAGAGG	GCATCTAG	CGCTGCTGC	CGCGCATCTG	GAACATCCGACAGGCACCAAG	67	
QY	1363	GAAAGCTCCTGGG	CGATCTGT	AAATAAACCTTTT	TTTCTTTGTTT	TTTAAAAAAA	1422	
Db	66	GAAAGCTCCTGGG	CGATCTGT	AAATAAACCTTTT	TTTCTTTGTTT	TTTAAAAAAA	7	
QY	1423	AAAAAA	1428					
Db	6	AAAAAA	1					

JD 6 AAAAAA 1 .

RESULT 8
PVT09714

BY708714	1014 bp	mRNA	linear	EST 16-DEC-2002
LOCUS				
BY708714				

DEFINITION
BY708714 RIKEN full-length enriched adult male stomach Mus musculus CDNA clone 2210403N03 5', mRNA sequence.

ACCESSION BY708714
VERSION BY708714.1
GI:27119906
INSTRUMENT CUNA C101E 4210703NOV 2 / INDIRA SEQUENCER.

VERSION BY708/14.1 GI:2/119906
 KEYWORDS
 SOURCE
 EST.
 Mus musculus (house mouse)
 SOURCE

SOURCE ORGANISM	Source
<i>Mus musculus</i> (house mouse)	1
<i>Mus musculus</i>	2
<i>Elanus caeruleus</i>	3
<i>Metazoa: Chordata: Vertebrata: Euteleostomi: Euteleostomi</i>	4

Eukaryota; Metazoa; Choriata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 (Pages 1 to 1014)
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,

ADAMS
Ozawa, I., Fukuoka, M., Asanuma, I., Nagai, C., Doi, N., Naito, S.,
Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamana, I.,
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,

Kiyosawa, H., Yagi, K., Iomaru, I., Hasegawa, Y., Nogami, A.,
Schonbach, C., Gojobori, T., Baldarelli, L., Hill, D. P., Bult, C.,
Hume, D. A., Quackenbush, J., Schriml, L. M., Kaneko, A., Matsuda, H.,

Hume, D.A., Quackenbush, J., Schriml, L.M., Kanpin, A., Matsuda, H.,
Batalov, S., Beisel, K.W., Blake, J.A., Bredt, D., Brusic, V.,
Chabot, J.P., Conning, C., Della E., Dierker, T.A.

Chothia, C., Cordani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,

Garbaldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. U., Jarvis, E. D., Kanai, A.,

Gubshinich, S., Hirokawa, N., Jackson, I. M., Kodav, S. D., Nakai, A., Kawai, H., Kawasaki, Y., Kedzierski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R.,

Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Magjoric, D. R.,

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 446)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

TITLE

Tumor Gene Index

JOURNAL

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

COMMENT

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40up from Gibco.

FEATURES

Location/Qualifiers

```
1..446
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2520674"
/tissue_type="colonic mucosa from 3 patients with Crohn's disease"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Soares-Dieckgraefe_colon_NHCD"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGGCGCGCGCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Tissue samples provided by Dr. Brian Dieckgraefe (Washington University, dieck@im.wustl.edu); colonic mucosa represents a range of disease; samples include both perforating (fistulas) and non-perforating samples. Library constructed by Bento Soares and W. Fatima Bonaldo."
```

ORIGIN

Query Match

31.0%; Score 446; DB 9; Length 446;
Best Local Similarity 100.0%; Pred. No. 6.1e-51;
Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 971 GCTCTCCGCGCTACCACTAGTACGAGCGCGCTTCTGGTCAAGTGGCAACCGCG 1030
DB 446 GCTCTCCGCGCTACCACTAGTACGAGCGCGCTTCTGGTCAAGTGGCAACCGCG 387
QY 1031 TCTGTGCTCTTCTCCGAGGCGCGCTGTAGTCTCCAGTATTTGCGCCGACGCTC 1090
DB 386 TCTGTGCTCTTCTCCGAGGCGCGGTGTAGTCTCCAGTATTTGCGCCGACGCTC 327
QY 1091 TTGCAACCTTCTGACCAAGGCCCAAGACTGACGAGGAGAGGAGGAGGAGCTC 1150
DB 326 TTGCAACCTTCTGACCAAGGCCCAAGACTGACGAGGAGAGGAGGAGGAGCTC 267
QY 1151 TTATCTCTGCGCAACCACTGACAGAGCGCGCTCTCCAGACTTAAATATGTATCA 1210
DB 266 TTATCTCTGCGCAACCACTGACAGAGCGCGCTCTCCAGACTTAAATATGTATCA 207
QY 1211 CTAACTGTAGAGGGGAGCCCAATCTGACTCTCCGCGCTTGGGAGCATGACGAGCG 1270
DB 206 CTAACTGTAGAGGGGAGCCCAATCTGACTCTCCGCGCTTGGGAGCATGACGAGCG 147
QY 1271 GAAGCAAGTCCCGCCAGGCGCTGGGCGAGAGCTTCAGAAAGGAGGAGGAGGAGCT 1330
DB 146 GAAGCAAGTCCCGCCAGGCGCTGGGCGAGAGAGCTTCAGAAAGGAGGAGGAGGAG 87
QY 1331 GCGCGAGGCGCTCGGAGCATCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1390
DB 86 GCGCGAGGCGCTCGGAGCATCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 27
QY 1391 AACCTTTTCTTTTGTATTTTAA 1416
DB 26 AACCTTTTCTTTTGTATTTTAA 1
```

RESULT 10

AA573825 466 bp mRNA linear EST 12-SEP-1997
AA573825/c
NC08a11.s1 NCI_CGAP_C02 Homo sapiens cDNA clone IMAGE:1012892 3',

DEFINITION

AA573825 466 bp mRNA linear EST 12-SEP-1997
AA573825/c
NC08a11.s1 NCI_CGAP_C02 Homo sapiens cDNA clone IMAGE:1012892 3',

ACCESSION

AA573825
AA573825
AA573825.1 GI:2348340

VERSION

AA573825
AA573825
AA573825.1 GI:2348340

KEYWORDS

EST.
Homo sapiens (human)

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

1 (bases 1 to 466)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

TITLE

Tumor Gene Index

JOURNAL

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

COMMENT

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40m3 fwd. RT from Amersham

FEATURES

Location/Qualifiers

```
1..466
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1012892"
/tissue_type="tumor"
/lab_host="SOIR (kanamycin resistant)"
/clone_lib="NCI_CGAP_C02"
/note="Organ: colon; Vector: Bluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: 01igo dt. Bulk colon villosus adenoma. 5' adaptor sequence: 5' GAATTCGGACGAG 3' 3' adaptor sequence: 5' CTCGACGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."
```

ORIGIN

Query Match

30.9%; Score 445.4; DB 9; Length 466;
Best Local Similarity 98.5%; Pred. No. 7.1e-51;
Matches 460; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

```
QY 947 TGCCGCTGCGCGCGCTCCGCTAGGCTCTCCGCGCTACACCTCAGTACGAGCGCGCT 1006
DB 466 TGCCGCTGCGCGCGCTCCGCTAGGCTCTCTCCGCGCTACACCTCAGTACGAGCGCGCT 407
QY 1007 TCTGGGTACGCTGCGCAACCGCGCTCTGTGCTCTTCTCTGAGAGGCGCGTGTGATC 1066
DB 406 TCTGGGTACGCTGCGCAACCGCGCGCTCTGTGCTCTTCTCTGAGAGGCGCGTGTGATC 347
QY 1067 TCCAGTATTTGGCGCCAGCGCTCTTGACACCTTTTGACCAAGGCGCCAGAGCTGCA 1126
DB 346 TCCAGTATTTGGCGCCAGCGCTCTTGACACCTTTTGACCAAGGCGCCAGAGCTGCA 287
QY 1127 GCGAGAGAGAGGAGGAGCTCACTTTATCTCTGCGGAGCCACTGACAGAGGCGCGCTC 1186
DB 286 GCGAGAGAGAGGAGGAGCTCACTTTATCTCTGCGGAGCCACTGACAGAGGCGCGCTC 227
QY 1187 TCCAGACTTAAATATGTATCACCACTAATCTGTAGAGGAGGAGCCCAATCTGACTCTCC 1246
DB 226 TCCAGACTTAAATATGTATCACCACTAATCTGTAGAGGAGGAGCCCAATCTGACTCTCC 167
```

QY 1247 CCAGCTTGGGACATCGACAGCCGAGGAGCAAGTGCCCGCAGAGCTGAGCCAGAGAGACTC 1306
 Db 166 CCGCTTGGGACATCGACAGCCGAGGAGCAAGTGCCCGCAGAGCTGAGCCAGAGAGACTC 107
 QY 1307 CAGAAAGGACACTGAGCCCTGCTGCGGAGAGCTCGGAATCCGAGGAGACAGAGGAAA 1366
 Db 106 CAGAAAGGACACTGAGCCCTGCTGCGGAGAGCTCGGAATCCGAGGAGACAGAGGAAA 48
 QY 1367 GTCTCTGGGAGCATCTGTAAATAACCTTTTCTTTTGTGTTTTT 1413
 Db 47 GTCTCTGGGAGCATCTGTAAATAACCTTTTCTTTTGTGTTTTT 1

RESULT 11
 A1991272 434 bp mRNA linear EST 09-MAR-2000
 A1991272/c wu4h04.x1 Soares Dieckgraefe.colon_NHCD Homo sapiens cDNA clone
 DEFINITION IMAGE:2522647 3', mRNA sequence.
 A1991272
 ACCESSION A1991272.1 GI:5838177
 VERSION EST.
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 TITLE Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Email: cgapbs-remail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 558 Std Error: 0.00
 Seg primer: -40UP from Gdbco.
 Location/Qualifiers

FEATURES
 source
 1..434
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2522647"
 /tissue_type="colonic mucosa from 3 patients with Crohn's
 disease"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="Soares Dieckgraefe.colon_NHCD"
 /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
 strand cDNA was primed with a Not I oligo(dT) primer [5'
 TGTTACCACTGAGAGTGGAGCGCGCGCTTTTCTTTTCTTTT 3']
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization. Tissue samples
 provided by Dr. Brian Dieckgraefe (Washington University,
 dieck@im.wustl.edu); colonic mucosa represents a range of
 disease involvement from moderate to severe Crohn's
 disease; samples include both perforating (fistulas) and
 non-perforating samples. Library constructed by Bento
 Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 30.1%; Score 434; DB 9; Length 434;
 Best Local Similarity 100.0%; Freq. No. 2.6e-49;
 Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 981 GCTCACTCACTAGTACGCGCGCTTGTGGTCAAGCTGCAACCGCGCTGTGCTCT 1040
 Db 434 GCTCACTCACTAGTACGCGCGCTTGTGGTCAAGCTGCAACCGCGCTGTGCTCT 375
 QY 1041 CTCTCTGAGAGGCGCGCTGTGAGTCTCCAGTATTTGGGCCAGCGCTCTTGCAACCT 1100
 Db 374 CTCTCTGAGAGGCGCGCTGTGAGTCTCCAGTATTTGGGCCAGCGCTCTTGCAACCT 315

QY 1101 TCTGACCAAAAGCCCAAGAGACTGACCCAGAGAGAGAGGGGCTCACTTTATCTCGG 1160
 Db 314 TCTGACCAAAAGCCCAAGAGACTGACCCAGAGAGAGAGGGGCTCACTTTATCTCGG 255
 QY 1161 CGACCCATCTGACAAAGAGCGCTCTCCGACACTTAAATGTATCAACCACTAACCTGTG 1220
 Db 254 CGACCCATCTGACAAAGAGCGCTCTCCGACACTTAAATGTATCAACCACTAACCTGTG 195
 QY 1221 AGGGGAGCCCAATCTGACTCTTCCCGCTTGGGACATCGAGCCCGGAGAGAGCTGC 1280
 Db 194 AGGGGAGCCCAATCTGACTCTTCCCGCTTGGGACATCGAGCCCGGAGAGAGCTGC 135
 QY 1281 CCGCCAGCCTTGGGCGCCAGAGAGACTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1340
 Db 134 CCGCCAGCCTTGGGCGCCAGAGAGACTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 75
 QY 1341 TCGACATCTCGACAGGACACAGGAAAGTCTCTGAGGCGCATCTGTAAATAACCTTTT 1400
 Db 74 TCGACATCTCGACAGGACACAGGAAAGTCTCTGAGGCGCATCTGTAAATAACCTTTT 15
 QY 1401 TCTTTTGTGTTTTT 1414
 Db 14 TCTTTTGTGTTTTT 1

RESULT 12
 AA593860/c 414 bp mRNA linear EST 25-SEP-1997
 LOCUS aa19103.ai NCI-CGAP_C012 Homo sapiens cDNA clone IMAGE:1084349 3',
 DEFINITION mRNA sequence.
 AA593860
 ACCESSION AA593860.1 GI:2408538
 VERSION EST.
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 TITLE Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Email: cgapbs-remail.nih.gov
 Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: Stratagene, Inc.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 www.bio.lnl.gov/dbtp/image/image.html
 Insert Length: 1204 Std Error: 0.00
 Seg primer: -40m13 fwd. EF from Amersham
 High quality sequence stop: 406.
 Location/Qualifiers

FEATURES

source

1..414
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1084349"
 /sex="mixed"
 /tissue_type="colon tumor"
 /lab_host="SQR (kanamycin resistant)"
 /clone_lib="NCI-CGAP_C012"
 /note="Organ: colon; Vector: Bluescript SK-; Site 1:
 EcoRI; Site 2: XhoI; cloned unidirectionally. Primer:
 Oligo dt. Pooled colon tumors, 5' adaptor sequence: 5'
 GAATTCGACAGAG 3' 3' adaptor sequence: 5'
 CTCAGATTTTTTTTTTTTTTTTTT 3' Average insert size: 1.2 kb."

ORIGIN

Query Match 28.7%; Score 414; DB 9; Length 414;
 Best Local Similarity 100.0%; Pred. No. 1.3e-46;
 Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 999 CGCCGCTCTTGGGTCACGCTGGGAACCGGCGCTCTTGGCTCTTCTTGGAGGGGCGCT 1058
 DB 414 CGCCGCTCTTGGGTCACGCTGGGAACCGGCGCTCTTGGCTCTTCTTGGAGGGGCGCT 355
 OY 1059 GGTGAGTCTCCAGATATGTTGGCCGACGCTCTTGGACCCCTTGGACCAAGCCCA 1118
 DB 354 GGTGAGTCTCCAGATATGTTGGCCGACGCTCTTGGACCCCTTGGACCAAGCCCA 295
 OY 1119 GCACTGACGACGAGAGAGAGGGGCTCACTCTTATCTTGGGACCCACTGACAGCA 1178
 DB 294 GCACTGACGACGAGAGAGAGGGGCTCACTCTTATCTTGGGACCCACTGACAGCA 235
 OY 1179 GCGCGCTCTCCAGACTTAAATGTATACCACTTACCTGAGGGGAGACCAATCTGA 1238
 DB 234 GCGCGCTCTCCAGACTTAAATGTATACCACTTACCTGAGGGGAGACCAATCTGA 175
 OY 1239 CTCCTTCCCGCTTGGGACATCGAGGCGCGGAGAGCATGCCCCGACGCTGGGCGAG 1298
 DB 174 CTCCTTCCCGCTTGGGACATCGAGGCGCGGAGAGCATGCCCCGACGCTGGGCGAG 115
 OY 1299 GAGAGCTCCAGAGAGGAGCATGAGGCTGAGGGGAGGAGCTGAGACATCGAGAGGAC 1358
 DB 114 GAGAGCTCCAGAGAGGAGCATGAGGCTGAGGGGAGGAGCTGAGACATCGAGAGGAC 55
 OY 1359 CAGGAGAACTCTCTGGGCGCATCTGTAATAAATCTTTTCTTTTCTTTT 1412
 DB 54 CAGGAGAACTCTCTGGGCGCATCTGTAATAAATCTTTTCTTTTCTTTT 1

RESULT 13
 A1983793/c 411 bp mRNA linear EST 27-OCT-1999
 LOCUS w120c09.x1 Soares Dieckgraefe.colon_NHCD Homo sapiens cDNA clone
 DEFINITION IMAGE:2520592.3, mRNA sequence.
 A1983793
 ACCESSION A1983793.1 GI:5811012
 VERSION EST.
 KEYWORDS Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 411)
 AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 This clone is available royalty-free through LNLN; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Seg primer: -40UP from Gidco
 High quality sequence stop: 400.
 Location/Qualifiers
 1..411
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2520592"
 /tissue_type="colonic mucosa from 3 patients with Crohn's
 disease"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="Soares Dieckgraefe.colon_NHCD"
 /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5']
 TGTTCACATCTGAAGTGGAGCGCGCTTTTCTTTTCTTTTCTTTT 3',
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library

ORIGIN

Query Match 28.4%; Score 409.4; DB 9; Length 411;
 Best Local Similarity 99.8%; Pred. No. 5.6e-46;
 Matches 410; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1008 CTGGGTACGCTGCAACCGCGCTCTGCTCTTCTTCTTGGAGGGGCGCTGGTGTCT 1067
 DB 411 CTGGGTACGCTGCAACCGCGCTCTGCTCTTCTTCTTGGAGGGGCGCTGGTGTCT 352
 OY 1068 CCAATATGTTGGCCGACGAGGCTCTTGGACCCCTTGGACCAAGGCGCAAGACTGGAG 1127
 DB 351 CCAATATGTTGGCCGACGAGGCTCTTGGACCCCTTGGACCAAGGCGCAAGACTGGAG 292
 OY 1128 CCAAGAGAGAGGGGCTCACTCTTATCTCTGGGACCACTGACCAAGAGCGCTCT 1187
 DB 291 CCAAGAGAGAGGGGCTCACTCTTATCTCTGGGACCACTGACCAAGAGCGCTCT 232
 OY 1188 CCAAGACTTAAATGTATACCACTTACCTGAGGGGAGACCAATCTGACTCTTCTTCC 1247
 DB 231 CCAAGACTTAAATGTATACCACTTACCTGAGGGGAGACCAATCTGACTCTTCTTCC 172
 OY 1248 CGCCTTGGGACATCGAGGCGGAGAGAGAGTGGCGCGAGGCTGGGCGAGAGAGCTCC 1307
 DB 171 CGCCTTGGGACATCGAGGCGGAGAGAGAGTGGCGCGAGGCTGGGCGAGAGAGCTCC 112
 OY 1308 AGAAGAGGACATGAGGCTGCTGCTGGCGAGGCTCTGACATCCGAGGACCGAGGAAG 1367
 DB 111 AGAAGAGGACATGAGGCTGCTGCTGGCGAGGCTCTGACATCCGAGGACCGAGGAAG 52
 OY 1368 TCTCTTGGGCGCATCTGTAATAAATCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1418
 DB 51 TCTCTTGGGCGCATCTGTAATAAATCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1

went through one round of normalization. Tissue samples
 provided by Dr. Brian Dieckgraefe (Washington University,
 dieckgrm.wustl.edu); colonic mucosa represents a range of
 disease involvement from moderate to severe Crohn's
 disease; samples include both perforating (fistulas) and
 non-perforating samples. Library constructed by Bento
 Soares and M. Fatima Bonaldi.

RESULT 14
 A1274929/c 405 bp mRNA linear EST 29-JAN-1999
 LOCUS g149c11.x1 NCI CGAP C08 Homo sapiens cDNA clone IMAGE:1875668.3
 DEFINITION similar to contains TAR1.t3 TAR1 repetitive element; mRNA
 sequence.
 A1274929
 ACCESSION A1274929.1 GI:3897203
 VERSION EST.
 KEYWORDS Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 405)
 AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Christopher Moskalkuk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CCAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
 www.bio.lnl.gov/bdrp/image/image.html
 Insert Length: 1458 Std Error: 0.00
 Seg primer: -40UP from Gidco
 High quality sequence stop: 395.

FEATURES

source

Location/Qualifiers

1. .405
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1875668"
 /issue_type="adenocarcinoma"
 /lab_host="DH10B"
 /lab_lib="NCI CGAP Co8"
 /note="Organ: Colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from colon adenocarcinoma, and was then primed with a Not I - oligo (dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match

28.1%; Score 405; DB 9; Length 405;

Best Local Similarity 100.0%; Pred. No. 2.2e-45;

Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1009 TGGGTACGCTGGCAACCGCGCTGCTGCTCTTCTTCGAGGGGCCGTGAGTCTC 1068
 Db 405 TGGGTACGCTGGCAACCGCGCTGCTGCTCTTCTTCGAGGGGCCGTGAGTCTC 346
 QY 1069 CAGTATGTTGGGCCGCGCTCTTGGACCCCTTTCGACCAAGGCCCAAGACTGCAGC 1128
 Db 345 CAGTATGTTGGGCCGCGCTCTTGGACCCCTTTCGACCAAGGCCCAAGACTGCAGC 286
 QY 1129 CAGAGAGAGGGGGCTCACCTCTTATCTCTGGCGACCACTGCAACAGAGCCGCTCTC 1188
 Db 285 CAGAGAGAGGGGGCTCACCTCTTATCTCTGGCGACCACTGCAACAGAGCCGCTCTC 226
 QY 1189 CCGACTTAAATGATATCACTTAACCTGTGAGGGGACCCAACTGTGACTCTTCTCC 1248
 Db 225 CCGACTTAAATGATATCACTTAACCTGTGAGGGGACCCAACTGTGACTCTTCTCC 166
 QY 1249 GCCTTGGGACATCGAGCGCGGGAAGCACTGCGCGGAGCCCTGAGGAGAGACTCA 1308
 Db 165 GCCTTGGGACATCGAGCGCGGGAAGCACTGCGCGGAGCCCTGAGGAGAGACTCA 106
 QY 1309 GGAAGGGCACTGAGCGCTGTGCGCGAGGCTCTGCAATCCGAGGACCAAGGAAAGT 1368
 Db 105 GGAAGGGCACTGAGCGCTGTGCGCGAGGCTCTGCAATCCGAGGACCAAGGAAAGT 46
 QY 1369 CTCTCTGGGGCGATCTGTAATAAACCCTTTTCTTTTCTTTT 1413
 Db 45 CTCTCTGGGGCGATCTGTAATAAACCCTTTTCTTTTCTTTT 1

RESULT 15

A1346155/c

405 bp mRNA linear EST 02-FEB-1999

LOCUS

DEFINITION gp43f12.x1 NCI CGAP Co8 Homo sapiens cDNA clone IMAGE:1875807 3' sequence.

ACCESSION A1346155

VERSION A1346155.1 GI:4083361

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 405)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

FEATURES

source

Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www.bio.lnlnl.gov/bdrrp/image/image.html
 Insert Length: 1488 Std Error: 0.00
 Seq primer: -40UP from Gldco
 High quality sequence stop: 375.

Location/Qualifiers

1. .405
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1925807"
 /issue_type="adenocarcinoma"
 /lab_host="DH10B"
 /clone_lib="NCI CGAP Co8"
 /note="Organ: Colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from colon adenocarcinoma, and was then primed with a Not I - oligo (dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match

27.8%; Score 400.2; DB 9; Length 405;

Best Local Similarity 99.3%; Pred. No. 9.9e-45;

Matches 402; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1009 TGGGTACGCTGGCAACCGCGCTGCTGCTCTTCTTCGAGGGGCCGTGAGTCTC 1068
 Db 405 TGGGTACGCTGGCAACCGCGCTGCTGCTCTTCTTCGAGGGGCCGTGAGTCTC 346
 QY 1069 CAGTATGTTGGGCCGCGCTCTTGGACCCCTTTCGACCAAGGCCCAAGACTGCAGC 1128
 Db 345 CAGTATGTTGGGCCGCGCTCTTGGACCCCTTTCGACCAAGGCCCAAGACTGCAGC 286
 QY 1129 CAGAGAGAGGGGGCTCACCTCTTATCTCTGGCGACCACTGCAACAGAGCCGCTCTC 1188
 Db 285 CAGAGAGAGGGGGCTCACCTCTTATCTCTGGCGACCACTGCAACAGAGCCGCTCTC 226
 QY 1189 CCGACTTAAATGATATCACTTAACCTGTGAGGGGACCCAACTGTGACTCTTCTCC 1248
 Db 225 CCGACTTAAATGATATCACTTAACCTGTGAGGGGACCCAACTGTGACTCTTCTCC 166
 QY 1249 GCCTTGGGACATCGAGCGCGGGAAGCACTGCGCGGAGCCCTGAGGAGAGACTCA 1308
 Db 165 GCCTTGGGACATCGAGCGCGGGAAGCACTGCGCGGAGCCCTGAGGAGAGACTCA 106
 QY 1309 GGAAGGGCACTGAGCGCTGTGCGCGAGGCTCTGCAATCCGAGGACCAAGGAAAGT 1368
 Db 105 GGAAGGGCACTGAGCGCTGTGCGCGAGGCTCTGCAATCCGAGGACCAAGGAAAGT 46
 QY 1369 CTCTCTGGGGCGATCTGTAATAAACCCTTTTCTTTTCTTTT 1413
 Db 45 CTCTCTGGGGCGATCAAAAATAAACCCTTTTCTTTTCTTTT 1

Search completed: February 23, 2004, 23:20:16
 Job time : 3671.04 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2004, 19:21:32, Search time 20 Seconds
(without alignments)
1539.064 Million cell updates/sec

Title: US-09-936-456-2
Perfect score: 320
Sequence: 1 MRLMNGVLFFYPQPRHAGF.....GDPHLKQALPDKCTITNL 320

Scoring table: OLIGO
Gapop 60.0, Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size: 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: PIR-78:*
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	2.5	141	2 AG0657	hypothetical perip
2	8	2.5	227	2 B83505	hypothetical prote
3	8	2.5	253	2 A49055	thioesterase - Str
4	8	2.5	417	2 JC7092	Paul protein - fis
5	8	2.5	575	2 B83586	probable permease
6	8	2.5	615	2 B91272	enzyme in methyl-d
7	8	2.5	615	2 B86113	enzyme in methyl-d
8	8	2.5	615	2 PH0853	methyl-directed mi
9	8	2.5	618	2 A33588	mismatch repair pr
10	8	2.5	618	2 AG1048	DNA mismatch repair
11	8	2.5	629	2 E64046	DNA mismatch repair
12	8	2.5	633	2 F83028	DNA mismatch repair
13	8	2.5	633	2 AC0046	DNA mismatch repair
14	8	2.5	653	2 AB2334	DNA mismatch repair
15	8	2.5	659	2 C83904	DNA mismatch repair
16	8	2.2	15	2 I46909	voltage-dependent
17	8	2.2	86	2 D91094	type III secretion
18	8	2.2	86	2 H85939	type III secretion
19	8	2.2	93	2 F95340	hypothetical prote
20	8	2.2	102	2 H72619	hypothetical prote
21	8	2.2	116	2 C86120	probable NADH-plas
22	8	2.2	116	2 D49339	probable growth in
23	8	2.2	116	2 C91279	cell growth regula
24	8	2.2	118	2 TC02049	probable growth in
25	8	2.2	126	2 F83294	lipid transfer pro
26	8	2.2	127	2 F83803	hypothetical prote
27	8	2.2	131	1 RDB215	hypothetical prote
28	8	2.2	182	2 S69469	hypothetical prote
29	8	2.2	205	1 F71193	hypothetical prote

30	7	2.2	205	2 G70338	conserved hypothet
31	7	2.2	206	2 D70935	hypothetical prote
32	7	2.2	209	2 E75152	hypothetical prote
33	7	2.2	210	2 AB7022	conserved hypothet
34	7	2.2	215	2 E83564	hypothetical prote
35	7	2.2	218	2 C81152	hypothetical prote
36	7	2.2	243	2 C69754	hypothetical prote
37	7	2.2	249	2 H84688	toxic cation resis
38	7	2.2	258	2 F90029	probable thaumatin
39	7	2.2	263	2 D95180	hypothetical prote
40	7	2.2	263	2 H98047	glutathione S-tran
41	7	2.2	266	2 E64955	conserved hypothet
42	7	2.2	266	2 C90961	filix protein precu
43	7	2.2	266	2 C85809	probable periplasm
44	7	2.2	266	2 AD0750	probable periplasm
45	7	2.2	266	2 T34172	cytine-binding pe
					hypothetical prote

ALIGNMENTS

RESULT 1
AG0657
hypothetical periplasmic protein STY1364 [imported] - Salmonella enterica subsp. ente
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AG0657
R:Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D., Main, J., Church
th, T., Connor, P., Cronin, A., Davis, P., Davies, R.M., Dowd, L., White, N., Farr
S., Moule, S., O'Garra, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens,
A., Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AG0657
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-141 <PAR-
A:Cross-references: GB:AL513382; P1DN:CAD01632.1; P1D:916502485; GSPDB:GN00176
C:Genetics:
A:Gene: STY1364

Query Match 2.5%; Score 8; DB 2; Length 141;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
DB 11 ALASISSV 18
|||||

RESULT 2
B83505
hypothetical protein PA1118 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: B83505
R:Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warren, P., Hickney, M.J.,
Adman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Laribig, K., L
J. Lory, S., Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: AB2950; MUID:20437337; PMID:10984043
A:Accession: B83505
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-227 <STO-
A:Cross-references: GB:AE004542; GB:AE004091; MUID:99947034; P1DN:AA04507.1; GSPDB:GN
A:Experimental source: Strain PA01
C:Genetics:
A:Gene: PA1118

A:Molecule type: DNA
A:Residues: 1-615 <STO>
A:Cross-references: GB:AE005174; NID:g1251915; PIRIN:AA059366.1; GSPDB:GN00145; UMGPP:A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics
A:Gene: mutL
C:Superfamily: mismatch repair protein hmxB

Query Match 2.5%; Score 8; DB 2; Length 615;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
 |||||
 Db 98 ALASISSV 105

RESULT 8
 PH0853
 methyl-directed mismatch repair protein mult. - Escherichia coli (strain K-12)
 C/Species: Escherichia coli
 C/Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 01-Mar-2002
 C/Accession: PH0853; S56395; A37318; S40056; E65227; S23011
 R/Tsai, H.C.T.; Mandavilli, B.S.; Winkler, M.E.
 Nucleic Acids Res. 20, 2379, 1992
 A/Title: Nonconserved segment of the mult. protein from Escherichia coli K-12 and Salmone
 A/Reference number: PH0853; MUID:92279041; PMID:1594459
 A/Accession: PH0853
 A/Molecule type: DNA
 A/Residues: 1-615 <TSU>
 A/Cross-references: EMBL:Z11831; NID:G42066; PIDN:CAA7850.1; PID:G42067
 A/Experimental source: strain K12
 A/Note: the complete nucleotide sequence is not shown; the complete translation is not s
 A/Note: translation of the nucleotide sequence is not complete
 R/Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
 Nucleic Acids Res. 23, 2105-2119, 1995
 A/Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.
 A/Reference number: S56314; MUID:95343362; PMID:7610040
 A/Accession: S56395
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-615 <BUR>
 A/Cross-references: EMBL:U14003; NID:G1263172; PIDN:AAA97066.1; PID:G537011
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994
 R/Connolly, D.M.; Winkler, M.E.
 J. Bacteriol. 173, 1711-1721, 1991
 A/Title: Structure of Escherichia coli K-12 *miaA* and characterization of the mutator phe
 A/Reference number: A37318; MUID:9154127; PMID:1959389
 A/Accession: A37318
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 596-615 <CON>
 A/Cross-references: GB:M63655; GB:M37459; NID:G146859; PIDN:AAA24173.1; PID:G146859
 R/Tsai, H.C.T.; Zhao, G.; Peng, G.; Leung, H.C.E.; Winkler, M.E.
 Mol. Microbiol. 11, 189-202, 1994
 A/Title: The mult. repair gene of Escherichia coli K-12 forms a superoperon with a gene e
 A/Reference number: S40053; MUID:94195106; PMID:7511774
 A/Accession: S40056
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-23 <TSW>
 A/Cross-references: EMBL:L19346; NID:G304911; PIDN:AAA20098.1; PID:G304915
 R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
 A.; Rose, D.J.; Mau, B.; Shaoh, Y.
 Science 277, 1453-1462, 1997
 A/Title: The complete genome sequence of Escherichia coli K-12.
 A/Reference number: A64720; MUID:97426617; PMID:9278503
 A/Accession: E65227
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-615 <BLAT>
 A/Cross-references: GB:AE000489; GB:U00096; NID:G1790607; PIDN:ACG7127.1; PID:G1790612;
 A/Experimental source: strain K-12, substrain MG1655
 C/Genetics:
 A/Gene: mutL
 A/Map position: 95 min
 C/Function:
 A/Description: this protein is required for methyl-directed mismatch repair in vivo and
 C/Superfamily: mismatch repair protein hexB
 C/Keywords: DNA repair

Query Match 2.5%; Score 8; DB 2; Length 615;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
 |||||
 Db 98 ALASISSV 105

RESULT 9
 A33588
 mismatch repair protein mult. - Salmoneella typhimurium
 C/Species: Salmoneella typhimurium
 C/Date: 27-Feb-1990 #sequence_revision 27-Feb-1990 #text_change 08-Oct-1999
 C/Accession: A33588
 R/Mankovich, J.A.; McIntyre, C.A.; Walker, G.C.
 J. Bacteriol. 171, 5325-5331, 1989
 A/Title: Nucleotide sequence of the Salmoneella typhimurium mult. gene required for mis
 A/Reference number: A33588; MUID:90008766; PMID:2676972
 A/Accession: A33588
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-618 <MAN>
 A/Cross-references: GB:M29687; NID:G154185; PIDN:AAA27166.1; PID:G154186
 C/Superfamily: mismatch repair protein hexB

Query Match 2.5%; Score 8; DB 2; Length 618;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
 |||||
 Db 98 ALASISSV 105

RESULT 10
 AG1048
 DNA mismatch repair protein [imported] - Salmoneella enterica subsp. enterica serovar
 C/Species: Salmoneella enterica subsp. enterica serovar Typh
 A/Note: this species has also been called Salmoneella typh
 C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
 C/Accession: AG1048
 R/Parikh, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
 ch, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr
 S.; Koule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
 A/Title: Complete genome sequence of a multiple drug resistant Salmoneella enterica se
 A/Reference number: AB0502; MUID:21534947; PMID:11677608
 A/Accession: AG1048
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-618 <PAR>
 A/Cross-references: GB:AL513382; PIDN:CAD06836.1; PID:G16505486; GSPDB:GN00176
 C/Genetics:
 A/Gene: mutL
 C/Superfamily: mismatch repair protein hexB

Query Match 2.5%; Score 8; DB 2; Length 618;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
 |||||
 Db 98 ALASISSV 105

RESULT 11
 B64046
 mismatch repair protein mult. - Haemophilus influenzae (strain Rd KW20)
 C/Species: Haemophilus influenzae

C/Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1995
 C/Accession: B64046
 R/Plasmid: R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirnness, E.F.; Kerlavage, A.; Gocayne, J.D.; Scott, J.; Shiley, R.; Liu, L.I.; Glodet, A.; Kelley, C.M.; Weidman, J.; D.M.; Brandon, R.C.; Fine, L.D.; Fritschman, J.L.; Furumasa, N.S.M.; Science 269, 496-512, 1995
 A/Authors: Ghem, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
 A/Reference number: A64000; MUID:9550630; PMID:7542800
 A/Accession: B64046
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-629 <TIGR>
 A/Cross-references: GB:U32692; GB:L42023; NID:g1573013; PIDN:AAC21745.1; PID:g1573016; T
 C/Genetics:
 A/Gene: mult
 C/Superfamily: mismatch repair protein hexB
 C/Keyword: DNA repair

Query Match 2.5%; Score 8; DB 2; Length 629;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISV 230
 |||||
 DB 98 ALASISV 105

RESULT 12

F83028
 DNA mismatch repair protein Mult PA4946 [imported] - Pseudomonas aeruginosa (strain PA01
 C/Species: Pseudomonas aeruginosa
 C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C/Accession: F83028
 R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.T.; B
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lam,
 .; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 A/Reference number: A82950; MUID:20437337; PMID:10964043
 A/Accession: F83028
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-633 <STO>
 A/Cross-references: GB:A6004907; GB:A6004091; NID:g9951217; PIDN:AAG08331.1; GSPDB:GN001
 A/Experimental source: strain PA01
 C/Genetics:
 A/Gene: mult; PA4946
 C/Superfamily: mismatch repair protein hexB

Query Match 2.5%; Score 8; DB 2; Length 633;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISV 230
 |||||
 DB 102 ALASISV 109

RESULT 13

AC0046
 DNA mismatch repair protein [imported] - Yersinia pestis (strain CO92)
 C/Species: Yersinia pestis
 C/Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
 C/Accession: AC0046
 R/Parhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.
 deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
 ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
 Nature 413, 523-527, 2001
 A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A/Reference number: AB0001; MUID:21470413; PMID:11586360
 A/Accession: AC0046
 A/Status: preliminary

A/Molecule type: DNA
 A/Residues: 1-635 <KOR>
 A/Cross-references: GB:AL590842; PIDN:CA0829230.1; PID:g15978469; GSPDB:GN00175
 C/Genetics:
 A/Gene: mult
 C/Superfamily: mismatch repair protein hexB

Query Match 2.5%; Score 8; DB 2; Length 635;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISV 230
 |||||
 DB 98 ALASISV 105

RESULT 14

A82334
 DNA mismatch repair protein Mult VC0345 [imported] - Vibrio cholerae (strain N16961
 C/Species: Vibrio cholerae
 C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
 C/Accession: A82334
 R/Heidelberg, J.F.; Eissen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R
 chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Seller
 l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
 A/Reference number: A82035; MUID:20406833; PMID:10952301
 A/Accession: A82334
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-653 <HEI>
 A/Cross-references: GB:A6004123; GB:A6003852; NID:g9654756; PIDN:AAF93518.1; GSPDB:G
 A/Experimental source: serogroup O1; strain N16961; biotype El Tor
 C/Genetics:
 A/Gene: VC0345
 A/Map position: 1
 C/Superfamily: mismatch repair protein hexB

Query Match 2.5%; Score 8; DB 2; Length 653;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISV 230
 |||||
 DB 98 ALASISV 105

RESULT 15

C89904
 DNA mismatch repair protein [imported] - Staphylococcus aureus (strain N315)
 C/Species: Staphylococcus aureus
 C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
 C/Accession: C89904
 R/Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.;
 ma, A.; Mizutani, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kato, C.; Sekimizu,
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
 Lancet 357, 1225-1240, 2001
 A/Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
 A/Reference number: A89756; MUID:21311952; PMID:11418146
 A/Accession: C89904
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-669 <KUR>
 A/Cross-references: GB:BA000018; PID:g13701096; PIDN:BA842391.1; GSPDB:GN00149
 A/Experimental source: strain N315
 C/Genetics:
 A/Gene: mult
 C/Superfamily: mismatch repair protein hexB

Query Match 2.5%; Score 8; DB 2; Length 669;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	223	ALASISSV	230
Db	99	ALASISSV	106

Search completed: February 23, 2004, 19:35:20
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2004, 19:32:12 ; Search time 23 Seconds
(without alignments)
718.274 Million cell updates/sec

Title: US-09-936-456-2

Perfect score: 320
Sequence: 1 MTLMNGVLPFYQPRHAGF.....GDPHQAALPDKCITTL 320

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/2/1aa/5A.COMB.pap:*
2: /cgn2_6/prodata/2/1aa/5B.COMB.pap:*
3: /cgn2_6/prodata/2/1aa/6A.COMB.pap:*
4: /cgn2_6/prodata/2/1aa/6B.COMB.pap:*
5: /cgn2_6/prodata/2/1aa/PCTUS.COMB.pap:*
6: /cgn2_6/prodata/2/1aa/backfiles1.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	2.5	227	4	US-09-252-991A-19468
2	8	2.5	326	4	US-09-134-001C-4934
3	8	2.5	361	2	US-08-209-521-5
4	8	2.5	361	3	US-08-961-810-1
5	8	2.5	361	3	US-08-352-902D-1
6	8	2.5	361	4	US-09-265-503B-1
7	8	2.5	602	4	US-09-252-991A-22527
8	8	2.5	615	3	US-08-676-444-44
9	8	2.5	659	4	US-09-252-991A-31126
10	8	2.5	674	4	US-09-543-681A-5715
11	7	2.2	91	4	US-09-369-247-83
12	7	2.2	118	1	US-08-207-904-4
13	7	2.2	118	1	US-08-207-904-19
14	7	2.2	142	4	US-09-489-039A-9418
15	7	2.2	144	4	US-09-252-991A-31261
16	7	2.2	152	4	US-09-543-681A-6097
17	7	2.2	158	4	US-09-489-039A-8515
18	7	2.2	169	4	US-09-328-352-6931
19	7	2.2	225	4	US-09-465-901-18
20	7	2.2	226	4	US-09-252-991A-27390
21	7	2.2	246	3	US-08-896-933-33
22	7	2.2	246	4	US-09-314-235-33
23	7	2.2	262	3	US-08-948-265-2
24	7	2.2	262	3	US-08-948-265-4
25	7	2.2	263	3	US-08-474-379C-42
26	7	2.2	263	3	US-09-146-249A-42
27	7	2.2	263	3	US-08-206-188B-42

28	7	2.2	264	4	US-09-252-991A-21359	Sequence 21359, A
29	7	2.2	265	1	US-07-688-352C-42	Sequence 42, Appl
30	7	2.2	265	4	US-09-252-991A-32803	Sequence 32803, A
31	7	2.2	276	4	US-09-543-681A-4338	Sequence 4338, Ap
32	7	2.2	287	4	US-09-328-352-6652	Sequence 6652, Ap
33	7	2.2	299	4	US-09-383-634-35	Sequence 35, Appl
34	7	2.2	300	4	US-09-489-039A-12250	Sequence 12250, A
35	7	2.2	311	4	US-09-134-001C-3846	Sequence 3846, Ap
36	7	2.2	313	3	US-08-926-842B-62	Sequence 62, Appl
37	7	2.2	315	4	US-09-489-039A-9274	Sequence 9274, Ap
38	7	2.2	336	2	US-08-474-379C-36	Sequence 36, Appl
39	7	2.2	336	3	US-09-146-249A-36	Sequence 36, Appl
40	7	2.2	336	3	US-08-206-188B-36	Sequence 36, Appl
41	7	2.2	341	1	US-08-311-731A-151	Sequence 151, Appl
42	7	2.2	386	1	US-07-688-352C-44	Sequence 44, Appl
43	7	2.2	386	5	PCT-US91-02714-41	Sequence 41, Appl
44	7	2.2	404	2	US-08-474-379C-44	Sequence 44, Appl
45	7	2.2	404	3	US-09-146-249A-44	Sequence 44, Appl

ALIGNMENTS

```

RESULT 1
US-09-252-991A-19468
Sequence 19468, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 19468
LENGTH: 227
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19468

Query Match      2.5% Score 8; DB 4; Length 227;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      98 RVGLLVGL 105
Db      105 RVGLLVGL 112

RESULT 2
US-09-134-001C-4934
Sequence 4934, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4934
LENGTH: 326
TYPE: PRT

```

/ ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4934

Query Match 2.5%; Score 8; DB 4; Length 36;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
DB 100 ALASISSV 107

RESULT 3
US-08-209-521-5
Sequence 5, Application US/08209521

GENERAL INFORMATION:
APPLICANT: Liskay, Robert M.
APPLICANT: Bronner, C. Eric
APPLICANT: Baker, Sean M.
APPLICANT: Bollag, Roni J.
APPLICANT: Kolodner, Richard D.
TITLE OF INVENTION: MAMMALIAN DNA MISMATCH REPAIR GENES
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kolisch, Hartwell, Dickinson, McCormack &
Heuser
STREET: 520 S.W. Yamhill, Suite 200
CITY: Portland
STATE: Oregon
COUNTRY: US
ZIP: 97204

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/209,521
FILING DATE: 08-MAR-1994
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Van Rysselberghe, Pierre C.
REGISTRATION NUMBER: 33,557
REFERENCE/DOCKET NUMBER: OHSU 306A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (503) 224-6655
TELEFAX: (503) 295-6679
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 361 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-209-521-5

Query Match 2.5%; Score 8; DB 2; Length 361;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
DB 98 ALASISSV 105

RESULT 4
US-08-961-810-1
Sequence 1, Application US/08961810
Patent No. 6155713
GENERAL INFORMATION:
APPLICANT: Liskay, Robert M.

APPLICANT: Bronner, C. Eric
APPLICANT: Baker, Sean M.
APPLICANT: Bollag, Roni J.
APPLICANT: Kolodner, Richard D.
TITLE OF INVENTION: COMPOSITIONS AND METHODS RELATING TO DNA
MISMATCH REPAIR GENES
NUMBER OF SEQUENCES: 134
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kolisch, Hartwell, Dickinson, McCormack &
Heuser
STREET: 520 S.W. Yamhill Street, Suite 200
CITY: Portland
STATE: Oregon
COUNTRY: U.S.A.
ZIP: 97204
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,810
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Van Rysselberghe, Pierre C.
REGISTRATION NUMBER: 33,557
REFERENCE/DOCKET NUMBER: OHSU 306B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (503) 224-6655
TELEFAX: (503) 295-6679
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 361 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA

US-08-961-810-1

Query Match 2.5%; Score 8; DB 3; Length 361;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
DB 98 ALASISSV 105

RESULT 5
US-08-352-902D-1
Sequence 1, Application US/08352902D
Patent No. 6191268
GENERAL INFORMATION:

APPLICANT: Liskay, Robert M.
APPLICANT: Bronner, C. Eric
APPLICANT: Baker, Sean M.
APPLICANT: Bollag, Roni J.
APPLICANT: Kolodner, Richard D.
TITLE OF INVENTION: COMPOSITIONS AND METHODS RELATING TO DNA
MISMATCH REPAIR GENES
NUMBER OF SEQUENCES: 149
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kolisch, Hartwell, Dickinson, McCormack &
Heuser
STREET: 520 S.W. Yamhill Street, Suite 200
CITY: Portland
STATE: Oregon
COUNTRY: U.S.A.
ZIP: 97204
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```
/
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/352,902D
/ FILING DATE: 09-Dec-1994
/ CLASSIFICATION: <Unknown>
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Van Rysselberghe, Pierre C.
/ REGISTRATION NUMBER: 33,557
/ REFERENCE/DOCKET NUMBER: OHSU 306B
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (503) 224-6655
/ TELEFAX: (503) 295-6679
/
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 361 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: CDNA
/ SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-08-352-902D-1

Query Match      2.5%; Score 8; DB 3; Length 361;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      223 ALA1SSSV 230
Db      98 ALA1SSSV 105

RESULT 6
US-09-265-503B-1
/ Sequence 1, Application US/09265503B
/ Patent No. 6538108
/ GENERAL INFORMATION:
/ APPLICANT: Liskey, Robert M.
/ APPLICANT: Bromer, C. Eric
/ APPLICANT: Baker, Sean M.
/ APPLICANT: Bollag, Roni J.
/ APPLICANT: Kolodner, Richard D.
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS
/ TITLE OF INVENTION: RELATING TO DNA MISMATCH REPAIR GENES
/ NUMBER OF SEQUENCES: 148
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Kolisch, Hartwell, Dickinson, McCormack & Heuser
/ STREET: 520 S.W. Yamhill Street, Suite 200
/ CITY: Portland
/ STATE: Oregon
/ COUNTRY: U.S.A.
/ ZIP: 97204
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/265,503B
/ FILING DATE: March 10, 1999
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Van Rysselberghe, Pierre C.
/ REGISTRATION NUMBER: 33,557
/ REFERENCE/DOCKET NUMBER: OHSU 306D
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (503) 224-6655
/ TELEFAX: (503) 295-6679
/ TELEX: 360619
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
```

```
/
/ LENGTH: 361 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: CDNA
US-09-265-503B-1

Query Match      2.5%; Score 8; DB 4; Length 361;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      223 ALA1SSSV 230
Db      98 ALA1SSSV 105

RESULT 7
US-09-252-991A-22527
/ Sequence 22527, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: 107196,136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,150
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 22527
/ LENGTH: 602
/ TYPE: PRT
/ ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22527

Query Match      2.5%; Score 8; DB 4; Length 602;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      247 GAAFWYTL 254
Db      343 GAAFWYTL 350

RESULT 8
US-08-676-444-44
/ Sequence 44, Application US/08676444A
/ Patent No. 6294325
/ GENERAL INFORMATION:
/ APPLICANT: Wetmur, James G.
/ TITLE OF INVENTION: CLONING AND EXPRESSION OF THERMOSTABLE
/ TITLE OF INVENTION: MOTL GENES AND PROTEINS AND USBS THEREFOR
/ FILE REFERENCE: MSW95-02
/ CURRENT APPLICATION NUMBER: US/08/676,444A
/ CURRENT FILING DATE: 1996-07-05
/ NUMBER OF SEQ ID NOS: 48
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 44
/ LENGTH: 615
/ TYPE: PRT
/ ORGANISM: Escherichia coli
US-08-676-444-44

Query Match      2.5%; Score 8; DB 3; Length 615;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      223 ALA1SSSV 230
Db      98 ALA1SSSV 105
```


RESULT 9
US-09-252-991A-33126
Sequence 33126, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 33126
LENGTH: 659
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-33126

Query Match 2.5%; Score 8; DB 4; Length 659;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
Db 128 ALASISSV 135

RESULT 10
US-09-543-681A-5715
Sequence 5715, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543,681A
CURRENT FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 5715
LENGTH: 674
TYPE: PRT
ORGANISM: Proteus mirabilis
US-09-543-681A-5715

Query Match 2.5%; Score 8; DB 4; Length 674;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
Db 103 ALASISSV 110

RESULT 11
US-09-369-247-83
Sequence 83, Application US/09369247
Patent No. 6569992
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 44 Human Secreted Proteins
FILE REFERENCE: P2024P1
CURRENT APPLICATION NUMBER: US/09/369,247
CURRENT FILING DATE: 1999-08-05
EARLIER APPLICATION NUMBER: 60/074,118

EARLIER FILING DATE: 1998-02-09
EARLIER APPLICATION NUMBER: 60/074,157
EARLIER FILING DATE: 1998-02-09
EARLIER APPLICATION NUMBER: 60/074,137
EARLIER FILING DATE: 1998-02-09
EARLIER APPLICATION NUMBER: 60/074,341
EARLIER FILING DATE: 1998-02-09
EARLIER APPLICATION NUMBER: 60/074,141
EARLIER FILING DATE: 1998-02-09
NUMBER OF SEQ ID NOS: 172
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 83
LENGTH: 91
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (91)
OTHER INFORMATION: Xaa equals stop translation
US-09-369-247-83

Query Match 2.2%; Score 7; DB 4; Length 91;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 LVFLALA 35
Db 5 LVFLALA 11

RESULT 12
US-08-207-904-4
Sequence 4, Application US/08207904
Patent No. 5477002
GENERAL INFORMATION:
APPLICANT: Tuttle, AnnMarie
TITLE OF INVENTION: Anchor-Specific cDNA Sequences, Genomic
TITLE OF INVENTION: DNA Sequences and Recombinant DNA Sequences
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/207,904
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/908,242
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Lazar, Steven R.
REGISTRATION NUMBER: 32,618
REFERENCE/DOCKET NUMBER: CGC 1624
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8615
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 118 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-207-904-4

Query Match 2.2%; Score 7; DB 1; Length 118;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 LVFLALA 35
 |||||
 Db 5 LVFLALA 11

RESULT 13
 US-08-207-904-19
 ; Sequence 19; Application US/08207904

; Patent No. 5477002
 ; GENERAL INFORMATION:
 ; APPLICANT: Tuttle, AnnMarie
 ; APPLICANT: Crossland, Lyle D.
 ; TITLE OF INVENTION: Antigen-Specific cDNA Sequences, Genomic
 ; TITLE OF INVENTION: DNA Sequences and Recombinant DNA Sequences
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: CIBA-GEIGY Corporation
 ; STREET: 7 Skyline Drive
 ; CITY: Hawthorne
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10532

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/207,904
 ; FILING DATE:
 ; CLASSIFICATION: 800
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/908,242
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Lazar, Steven R.
 ; REGISTRATION NUMBER: 32,618
 ; REFERENCE/DOCKET NUMBER: CGC 1624
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (919)541-8615
 ; TELEFAX: (919)541-8689
 ; INFORMATION FOR SEQ ID NO: 19:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 118 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-207-904-19

Query Match 2.2%; Score 7; DB 1; Length 118;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 LVFLALA 35
 |||||
 Db 5 LVFLALA 11

RESULT 14
 US-09-489-039A-9418

; Sequence 9418; Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.2004001
 ; CURRENT APPLICATION NUMBER: US/09/489,039A

; CURRENT FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: US 60/117,747
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 9418
 ; LENGTH: 142
 ; TYPE: PRT
 ; ORGANISM: Klebsiella pneumoniae
 ; US-09-489-039A-9418

Query Match 2.2%; Score 7; DB 4; Length 142;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 VLLSLFI 63
 |||||
 Db 67 VLLSLFI 73

RESULT 15
 US-09-252-991A-31261
 ; Sequence 31261; Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 31261
 ; LENGTH: 144
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 ; US-09-252-991A-31261

Query Match 2.2%; Score 7; DB 4; Length 144;
 Best Local Similarity 100.0%; Pred. No. 64;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 FSAARVT 96
 |||||
 Db 106 FSAARVT 112

Search completed: February 23, 2004, 19:35:55
 Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2004, 19:33:22 ; Search time 36 seconds

(without alignments)
1876.918 Million cell updates/sec

Title: US-09-936-456-2

Perfect score: 320
Sequence: 1 MTLMNGVLPPYPOPRHAGF.....GDPHKQALPDLCITTNL 320

Scoring table:
Gapop 60.0 , Gapext 60.0

Searched: 809742 seqs, 21153259 residues

Word size : 0

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:

1: /cgn2_6/prodata/1/pubppa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/1/pubppa/PCT_NEW_PUB.pep.*
3: /cgn2_6/prodata/1/pubppa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/1/pubppa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/1/pubppa/US07_NEW_PUB.pep.*
6: /cgn2_6/prodata/1/pubppa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/prodata/1/pubppa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/1/pubppa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/1/pubppa/US09_PUBCOMB.pep.*
10: /cgn2_6/prodata/1/pubppa/US09C_PUBCOMB.pep.*
11: /cgn2_6/prodata/1/pubppa/US09C_PUBCOMB.pep.*
12: /cgn2_6/prodata/1/pubppa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata/1/pubppa/US10A_PUBCOMB.pep.*
14: /cgn2_6/prodata/1/pubppa/US10B_PUBCOMB.pep.*
15: /cgn2_6/prodata/1/pubppa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata/1/pubppa/US10_NEW_PUB.pep.*
17: /cgn2_6/prodata/1/pubppa/US60_NEW_PUB.pep.*
18: /cgn2_6/prodata/1/pubppa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	320	100.0	320	15 US-10-187-657-1	Sequence 1, Appl
2	219	68.4	320	15 US-10-264-237-2502	Sequence 2502, Ap
3	19	5.9	98	14 US-10-106-698-7554	Sequence 7554, Ap
4	19	5.9	112	10 US-09-759-130B-427	Sequence 427, App
5	19	5.9	112	13 US-10-042-431-57	Sequence 57, Appl
6	19	5.9	298	15 US-10-094-749-2834	Sequence 2834, Ap
7	19	5.9	343	10 US-09-759-130B-425	Sequence 425, Appl
8	19	5.9	343	13 US-10-042-431-55	Sequence 55, Appl
9	8	2.5	49	10 US-09-764-891-2983	Sequence 2983, Ap
10	8	2.5	120	14 US-10-156-761-12862	Sequence 12862, A
11	8	2.5	345	14 US-10-128-714-8106	Sequence 8106, Ap
12	8	2.5	615	14 US-10-109-791A-20	Sequence 20, Appl
13	8	2.5	629	9 US-09-815-242-10961	Sequence 10961, A
14	8	2.5	633	9 US-09-815-242-12077	Sequence 12077, A
15	8	2.5	669	9 US-09-815-242-5262	Sequence 5262, Ap

16	8	2.5	669	9 US-09-815-242-12334	Sequence 12334, A
17	8	2.5	669	14 US-10-109-791A-19	Sequence 19, Appl
18	8	2.5	734	9 US-09-738-626-4127	Sequence 4127, Ap
19	8	2.5	772	9 US-09-935-799A-2	Sequence 2, Appl
20	8	2.5	772	9 US-09-935-799A-5	Sequence 5, Appl
21	7	2.2	20	10 US-09-986-480-234	Sequence 234, App
22	7	2.2	33	14 US-10-081-816-89	Sequence 89, Appl
23	7	2.2	54	9 US-09-729-674-40	Sequence 40, Appl
24	7	2.2	54	9 US-09-925-301-1675	Sequence 1675, Ap
25	7	2.2	65	9 US-09-864-761-44598	Sequence 44598, A
26	7	2.2	87	11 US-09-864-408A-5048	Sequence 5048, Ap
27	7	2.2	91	14 US-10-062-548-833	Sequence 83, Appl
28	7	2.2	109	9 US-09-864-761-4256	Sequence 4256, A
29	7	2.2	126	10 US-09-820-843A-91	Sequence 91, Appl
30	7	2.2	147	15 US-10-108-260A-3104	Sequence 3104, Ap
31	7	2.2	194	15 US-10-108-260A-4172	Sequence 4172, Ap
32	7	2.2	206	14 US-10-080-170-513	Sequence 513, App
33	7	2.2	209	14 US-10-156-761-10982	Sequence 10982, A
34	7	2.2	210	14 US-10-080-170-999	Sequence 99, Appl
35	7	2.2	220	15 US-10-369-493-12396	Sequence 12396, A
36	7	2.2	225	14 US-10-259-430-18	Sequence 18, Appl
37	7	2.2	225	14 US-10-259-423-18	Sequence 18, Appl
38	7	2.2	225	14 US-09-770-075-2	Sequence 2, Appl
39	7	2.2	262	9 US-09-770-075-4	Sequence 4, Appl
40	7	2.2	262	9 US-09-769-997-2	Sequence 2, Appl
41	7	2.2	262	9 US-09-769-997-4	Sequence 4, Appl
42	7	2.2	262	14 US-10-136-761-14181	Sequence 14181, A
43	7	2.2	263	10 US-09-769-787-68	Sequence 68, Appl
44	7	2.2	266	9 US-09-815-242-10197	Sequence 10197, A
45	7	2.2	282	9 US-09-815-242-14057	Sequence 14057, A

ALIGNMENTS

RESULT 1
US-10-187-657-1
Sequence 1, Application US/10187657
Publication No. US20030068311A1
GENERAL INFORMATION:
APPLICANT: Lasek, Amy K.W.
APPLICANT: Baugm, Mariah R.
TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
FILE REFERENCE: PV-0009 CIP
CURRENT APPLICATION NUMBER: US/10/187,657
PRIOR FILING DATE: 2002-07-01
PRIOR APPLICATION NUMBER: PCT/US00/07817
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60/139,565
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PERL Program
SEQ ID NO 1
LENGTH: 320
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Inocyte ID No. US20030068311A1 4901066CD1
US-10-187-657-1
Query Match 100.0%; Score 320; DB 14; Length 320;
Best Local Similarity 100.0%; Pred. No. 4.86-293;
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Cy 1 MTLMNGVLPPYPOPRHAGFVSPLILVILFALASFLILPGIRGRSFMVLYRLS 60
Db 1 MTLMNGVLPPYPOPRHAGFVSPLILVILFALASFLILPGIRGRSFMVLYRLS 60
Cy 61 LFIGAIVAVHSAEMFVGTVTNTSYAFSAARVTARVGLVGLGGINITLTGTPTVQL 120
Db 61 LFIGAIVAVHSAEMFVGTVTNTSYAFSAARVTARVGLVGLGGINITLTGTPTVQL 120

```
QY 121 NETIDYNEQFTWRLKENYAAYANALEKGLDPVLVLAKEFTPSSPGLYHQYLAHGYA 180
DB 121 NETIDYNEQFTWRLKENYAAYANALEKGLDPVLVLAKEFTPSSPGLYHQYLAHGYA 180
QY 181 SATLWVAFCFWLISNVLLSTPAFLYGLALLTTGAFAFGFALASISSVPLCPRLGSS 240
DB 181 SATLWVAFCFWLISNVLLSTPAFLYGLALLTTGAFAFGFALASISSVPLCPRLGSS 240
QY 241 ALTTGYGAFWTLATGVLCFLGGAVVSLOQVYPSALRTLLDQSAKDCSQERGSSPLL 300
DB 241 ALTTGYGAFWTLATGVLCFLGGAVVSLOQVYPSALRTLLDQSAKDCSQERGSSPLL 300
QY 301 GDBLHQALPDLKCTITNL 320
DB 301 GDBLHQALPDLKCTITNL 320
```

RESULT 2

```
US-10-264-237-2502
; Sequence 2502, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
; APPLICANT: Btise et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA131P1
; CURRENT APPLICATION NUMBER: US/10/264,237
; PRIOR FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 2502
; LENGTH: 320
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (106)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-237-2502
```

Query Match 68.4%; Score 219; DB 15; Length 320;
Best Local Similarity 99.7%; Pred. No. 7.2e-198;
Matches 319; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 MTLMNGVLPFYPPORRAAGFSVPLIIVLIFLALASFLILPGIRGSRMFWLVRVLLS 60
DB 1 MTLMNGVLPFYPPORRAAGFSVPLIIVLIFLALASFLILPGIRGSRMFWLVRVLLS 60
QY 61 LFTGAELVAVHFAEAFVGTNTNTSYKAFSAARVARGLVLEG3INILTTGPHQOL 120
DB 61 LFTGAELVAVHFAEAFVGTNTNTSYKAFSAARVARGLVLEG3INILTTGPHQOL 120
QY 121 NETIDYNEQFTWRLKENYAAYANALEKGLDPVLVLAKEFTPSSPGLYHQYLAHGYA 180
DB 121 NETIDYNEQFTWRLKENYAAYANALEKGLDPVLVLAKEFTPSSPGLYHQYLAHGYA 180
QY 181 SATLWVAFCFWLISNVLLSTPAFLYGLALLTTGAFAFGFALASISSVPLCPRLGSS 240
DB 181 SATLWVAFCFWLISNVLLSTPAFLYGLALLTTGAFAFGFALASISSVPLCPRLGSS 240
QY 241 ALTTGYGAFWTLATGVLCFLGGAVVSLOQVYPSALRTLLDQSAKDCSQERGSSPLL 300
DB 241 ALTTGYGAFWTLATGVLCFLGGAVVSLOQVYPSALRTLLDQSAKDCSQERGSSPLL 300
QY 301 GDBLHQALPDLKCTITNL 320
DB 301 GDBLHQALPDLKCTITNL 320
```

RESULT 3

```
US-10-106-698-7554
; Sequence 7554, Application US/10106698
; Publication No. US20030103690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypep
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 7554
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (65)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC FEATURE
; LOCATION: (85)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC FEATURE
; LOCATION: (88)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC FEATURE
; LOCATION: (98)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-7554
```

Query Match 5.9%; Score 19; DB 14; Length 98;
Best Local Similarity 100.0%; Pred. No. 7.1e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 50 RWFMLVRVLLSLFTGAEIV 68
DB 8 RWFMLVRVLLSLFTGAEIV 26
```

RESULT 4

```
US-09-759-130B-427
; Sequence 427, Application US/09759130B
; Publication No. US2003002279A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: McCarthy, Sean A.
; APPLICANT: Fraser, Christopher C
; APPLICANT: Sharp, John D
; APPLICANT: Barnes, Thomas S
; APPLICANT: Kiser, Susan J
; APPLICANT: Mackay, Charles R
; APPLICANT: Meyers, Paul S
; APPLICANT: Leiby, Kevin R
; APPLICANT: Wrighton, Nicolas
; APPLICANT: Goddard, Andrew
; APPLICANT: Holtzman, Douglas A
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
; FILE REFERENCE: MP100-5350MIM
; CURRENT APPLICATION NUMBER: US/09/759,130B
; PRIOR FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 09/479,249
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/559,497
; PRIOR FILING DATE: 2000-04-27
```

```

; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/596,194
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/342,364
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/608,452
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/393,996
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: US 09/602,871
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/420,707
; PRIOR FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 427
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-759-130B-427
```

```
Query Match          5.9%; Score 19; DB 10; Length 112;
Best Local Similarity 100.0%; Pred. No. 8e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      145 ALEKGLDPVLYAEKFTP 163
Db      75 ALEKGLDPVLYAEKFTP 93
```

```

RESULT 5
US-10-042-431-57
; Sequence 57, Application US/10042431
; Publication No. US20020182675A1
; GENERAL INFORMATION:
; APPLICANT: MCCARTHY, Sean A
; APPLICANT: FRASER, Christopher C
; APPLICANT: SHARP, John D
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING DIAGNOSTIC,
; FILE REFERENCE: 10147-6U2
; CURRENT APPLICATION NUMBER: US/10/042,431
; PRIOR FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/578,063
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 57
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-042-431-57
```

```
Query Match          5.9%; Score 19; DB 13; Length 112;
Best Local Similarity 100.0%; Pred. No. 8e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      145 ALEKGLDPVLYAEKFTP 163
Db      75 ALEKGLDPVLYAEKFTP 93
```

```

RESULT 6
US-10-094-749-2834
; Sequence 2834, Application US/10094749
; Publication No. US2003021974A1
; GENERAL INFORMATION:
```

```

; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOTIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTYUKI
; APPLICANT: NAGAIARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 08435/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 2834
; LENGTH: 298
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2834
```

```
Query Match          5.9%; Score 19; DB 15; Length 298;
Best Local Similarity 100.0%; Pred. No. 2e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      145 ALEKGLDPVLYAEKFTP 163
Db      100 ALEKGLDPVLYAEKFTP 118
```

```

RESULT 7
US-09-759-130B-425
; Sequence 425, Application US/09759130B
; Publication No. US20030022279A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: McCarthy, Sean A
; APPLICANT: Fraser, Christopher C
; APPLICANT: Sharp, John D
; APPLICANT: Barnes, Thomas S
; APPLICANT: Kirt, Susan J
; APPLICANT: Mackay, Charles R
; APPLICANT: Myers, Paul S
; APPLICANT: Leiby, Kevin R
; APPLICANT: Wrighton, Nicolas
; APPLICANT: Goodearl, Andrew
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
; FILE REFERENCE: MFI00-5350N1M
; CURRENT APPLICATION NUMBER: US/09/759,130B
; PRIOR FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 09/479,249
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/559,497
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: US 09/333,159
```

```

; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/596,194
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/342,364
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/608,452
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/393,996
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: US 09/602,871
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/420,707
; PRIOR FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 425
; LENGTH: 343
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-759-1308-425

```

```

Query Match          5.9%; Score 19; DB 10; Length 343;
Best Local Similarity 100.0%; Pred.No. 2.3e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 145 ALEKGLDPVLYIAEKFTP 163
DB 145 ALEKGLDPVLYIAEKFTP 163

```

```

RESULT 8
US-10-042-431-55
; Sequence 55, Application US/10042431
; Publication No. US20020182675A1
; GENERAL INFORMATION:
; APPLICANT: MCCARTHY, Sean A
; APPLICANT: BARNES, Thomas M
; APPLICANT: FRASER, Christopher C
; APPLICANT: SHARP, John D
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING DIAGNOSTIC,
; FILE REFERENCE: 10147-602
; CURRENT APPLICATION NUMBER: US/10/042,431
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 55
; LENGTH: 343
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-042-431-55

```

```

Query Match          5.9%; Score 19; DB 13; Length 343;
Best Local Similarity 100.0%; Pred.No. 2.3e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 145 ALEKGLDPVLYIAEKFTP 163
DB 145 ALEKGLDPVLYIAEKFTP 163

```

```

RESULT 9
US-09-764-891-2983
; Sequence 2983, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006

```

```

; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 2983
; LENGTH: 49
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-891-2983

```

```

Query Match          2.5%; Score 8; DB 10; Length 49;
Best Local Similarity 100.0%; Pred.No. 8.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 55 VRVLSL 62
DB 39 VRVLSL 46

```

```

RESULT 10
US-10-156-761-12862
; Sequence 12862, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO: 12862
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12862

```

```

Query Match          2.5%; Score 8; DB 14; Length 120;
Best Local Similarity 100.0%; Pred.No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 263 LGGAVVSL 270
DB 34 LGGAVVSL 41

```

```

RESULT 11
US-10-128-714-8106
; Sequence 8106, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Weng
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Broshkin, Alexey M
; APPLICANT: Lemieux, Sebastian M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus an
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697

```

```

; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/235,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8106
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-8106

Query Match
Best Local Similarity 100.0%; Pred. No. 53;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 VRVLSLF 62
Db 46 VRVLSLF 53

RESULT 12
US-10-109-791A-20
; Sequence 20, Application US/10109791A
; Publication No. US20030138787A1
; GENERAL INFORMATION:
; APPLICANT: Biftech Oncologic Corp.
; TITLE OF INVENTION: Functional Genetic Tests of DNA Mismatch Repair
; FILE REFERENCE: BROL_102_NP
; CURRENT APPLICATION NUMBER: US/10/109,791A
; CURRENT FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 315
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 615
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-109-791A-20

Query Match
Best Local Similarity 100.0%; Pred. No. 90;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISV 230
Db 98 ALASISV 105

RESULT 13
US-09-815-242-10961
; Sequence 10961, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10961
; LENGTH: 629
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-815-242-10961

Query Match
Best Local Similarity 100.0%; Pred. No. 92;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISV 230
Db 98 ALASISV 105

RESULT 14
US-09-815-242-12077
; Sequence 12077, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12077
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-12077

Query Match
Best Local Similarity 100.0%; Pred. No. 93;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 223 ALASISSV 230
 |||||
 Db 102 ALASISSV 109

RESULT 15

US-09-815-242-5262
 ; Sequence 5262, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyckind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; TITLE OF INVENTION: Prokaryotes
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5262
 ; LENGTH: 669
 ; TYPE: PRF
 ; ORGANISM: Staphylococcus aureus
 US-09-815-242-5262

Query Match 2.5%; Score 8; DB 9; Length 669;
 Best Local Similarity 100.0%; Pred. No. 97;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
 |||||
 Db 99 ALASISSV 106

Search completed: February 23, 2004, 19:36:43
 Job time : 37 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2004, 19:20:36 ; Search time 45 Seconds
(without alignments)
2243.685 Million cell updates/sec

Title: US-09-936-456-2
Perfect score: 320
Sequence: 1 MTLNMGVLPFYPPQPRHAGF.....GDPHKQAALPDKCITNTL 320

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

SPREMBL 25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp Vertebrate:*
14: sp Unclassified:*
15: sp_rvlnus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	11.2	320	11 Q9D311	Q9D311 mus musculu
2	33	10.3	320	11 Q9D7U8	Q9D7U8 mus musculu
3	19	5.9	298	4 Q96MT4	Q96MT4 homo sapien
4	19	5.9	341	11 Q8YF49	Q8YF49 mus musculu
5	19	5.9	483	4 Q8N6K9	Q8N6K9 homo sapien
6	8	2.5	96	16 Q8XNQ5	Q8XNQ5 clostridium
7	8	2.5	120	16 Q82CL7	Q82CL7 streptomyce
8	8	2.5	141	16 Q827B8	Q827B8 salmonella
9	8	2.5	145	4 Q8N206	Q8N206 homo sapien
10	8	2.5	227	16 Q9T4L7	Q9T4L7 pseudomonas
11	8	2.5	253	2 Q54145	Q54145 streptomyce
12	8	2.5	338	16 Q83P96	Q83P96 shigella fl
13	8	2.5	382	5 Q7YTN3	Q7YTN3 caenorhabdi
14	8	2.5	417	3 Q9UR09	Q9UR09 schizosacch
15	8	2.5	468	2 Q9LAP6	Q9LAP6 staphylococ
16	8	2.5	504	16 Q9BE44	Q9BE44 rhizobium 1

17	8	2.5	507	16 Q8DMD2	Q8DMD2 synechococc
18	8	2.5	575	16 Q91642	Q91642 pseudomonas
19	8	2.5	598	16 Q7VN43	Q7VN43 haemophilus
20	8	2.5	615	16 Q7UJ40	Q7UJ40 shigella fl
21	8	2.5	616	16 Q83P95	Q83P95 shigella fl
22	8	2.5	631	16 Q8EJ70	Q8EJ70 shewanella
23	8	2.5	632	16 Q8BD01	Q8BD01 pseudomonas
24	8	2.5	645	16 Q8CPE9	Q8CPE9 staphylococ
25	8	2.5	645	16 Q87WJ2	Q87WJ2 pseudomonas
26	8	2.5	725	2 Q52978	Q52978 rhizobium m
27	8	2.5	734	16 Q8N5I6	Q8N5I6 corynebacte
28	8	2.5	850	16 Q89NM3	Q89NM3 bradyrhizob
29	8	2.5	926	2 Q8RTB9	Q8RTB9 rhizobium m
30	8	2.5	999	16 Q921V7	Q921V7 rhizobium m
31	7	2.2	15	6 Q28822	Q28822 oryctolagus
32	7	2.2	29	4 Q8TA15	Q8TA15 homo sapien
33	7	2.2	30	6 Q8MHW1	Q8MHW1 sus scrofa
34	7	2.2	56	6 Q8MHW9	Q8MHW9 sus scrofa
35	7	2.2	58	2 Q93NM4	Q93NM4 streptococc
36	7	2.2	72	16 Q7U7C7	Q7U7C7 synechococc
37	7	2.2	86	2 Q9E216	Q9E216 sodalis glo
38	7	2.2	86	2 Q93RC2	Q93RC2 escherichia
39	7	2.2	86	2 Q8GAT3	Q8GAT3 primary end
40	7	2.2	86	16 Q8X6P2	Q8X6P2 escherichia
41	7	2.2	93	16 Q92Z62	Q92Z62 rhizobium m
42	7	2.2	97	2 Q8EY73	Q8EY73 pseudomonas
43	7	2.2	100	16 Q93Y60	Q93Y60 streptococc
44	7	2.2	100	16 Q8NZ06	Q8NZ06 streptococc
45	7	2.2	102	17 Q9YC34	Q9YC34 aeropyrum p.

ALIGNMENTS

RESULT 1
ID Q9D311 PRELIMINARY; PRT; 320 AA.
AC Q9D311;
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TEMBLrel. 22, Last annotation update)
DE 9030623N16R1k protein (R1KEN CDNA 9030623N16 gene).
GN 9030623N16R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN
RP
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Colon;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shitagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanao I.,
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirral L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Blake J., Boffelli D., Bojunga N., Cammici P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Guentrich S., Hill D., Holman W., Home D.A., Kamaya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli U., Nombere P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyshewski A., Yoshida K., Hasegawa Y., Kawaji H., Kotsuki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.",
RL Nature 409:685-690(2001).
RN
RP
SEQUENCE FROM N.A.

RC TISSUE=Colon;
 RA Strusberg R.;
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK018569; BAB31281.1; -
 DR EMBL; BC011111; AAH31111.1; -
 DR MGD; MGI:1914061; 9030623N16R1K.
 SQ SEQUENCE 320 AA; 35427 MW; 4B7D4C81685162P7 CRC64;

Query Match 11.2%; Score 36; DB 11; Length 320;
 Best Local Similarity 100.0%; Pred. No. 6,7e-26;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 ALEKGLPDPVLYLAKEFTSPSCGLHYQYHAGHYA 180
 DB 145 ALEKGLPDPVLYLAKEFTSPSCGLHYQYHAGHYA 180

RESULT 2

ID Q9D7U8 PRELIMINARY; PRT; 320 AA.

AC Q9D7U8, 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE 9030623N16R1K protein.
 CN 9030623N16R1K.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCB1_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Stomach;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itch M., Ishii Y.,
 RA Aizawa K., Hara A., Fukunishi Y., Kondo H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl F., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gusninch S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Noridone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wymshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohetsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL; AK008816; BAB55910.1; -
 DR MGD; MGI:1914061; 9030623N16R1K.
 SQ SEQUENCE 320 AA; 36342 MW; 9541848376D0E2B2 CRC64;

Query Match 10.3%; Score 33; DB 11; Length 320;
 Best Local Similarity 100.0%; Pred. No. 5,2e-23;
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 ILPGIRGSRNFWLVRVLSLFGAETIVAVHFS 73
 DB 41 ILPGIRGSRNFWLVRVLSLFGAETIVAVHFS 73

RESULT 3

ID Q96M14 PRELIMINARY; PRT; 298 AA.

AC Q96M14, 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)

DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Hypothetical protein FLJ32334.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NC NCB1_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RC TISSUE=Prostate;
 RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
 RA Horita T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J., Chiba Y.,
 RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
 RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
 RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
 RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
 RA Magatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
 RA Suzuki Y., Sugano S., Nagahari K., Masuo Y., Nagai K., Isogai T.,
 RT "NEDD human cDNA sequencing project."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RA Strusberg R.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK056896; BAB71304.1; -
 DR EMBL; BC020811; AAH20841.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 298 AA; 33110 MW; D9B346C25D142E1D CRC64;

Query Match 5.9%; Score 19; DB 4; Length 298;
 Best Local Similarity 100.0%; Pred. No. 1,4e-09;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 ALEKGLPDPVLYLAKEFTP 163
 DB 100 ALEKGLPDPVLYLAKEFTP 118

RESULT 4

ID Q8VE49 PRELIMINARY; PRT; 341 AA.

AC Q8VE49, 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Hypothetical protein.
 CN BC019755.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCB1_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Strusberg R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC019755; AAH19755.1; -
 DR MGD; MGI:2384861; BC019755.
 KW Hypothetical protein.
 SQ SEQUENCE 341 AA; 37595 MW; CC1ED7AD103A7276 CRC64;

Query Match 5.9%; Score 19; DB 11; Length 341;
 Best Local Similarity 100.0%; Pred. No. 1,6e-09;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 ALEKGLPDPVLYLAKEFTP 163
 DB 145 ALEKGLPDPVLYLAKEFTP 163

RESULT 5

ID Q8N6K9 PRELIMINARY; PRT; 483 AA.

AC Q8N6K9, 01-DEC-2001 (TREMBlrel. 19, Last sequence update)

```

DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Similar to RIKEN CDNA 903063N16 gene.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, and Lung;
RA Strausberg R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC029819; AAB29819.1; -
SQ SEQUENCE 483 AA; 53519 MW; 54BFF64528B96C18 CRC64;

Query Match 5.9%; Score 19; DB 4; Length 483;
Best Local Similarity 100.0%; Pred. No. 2.2e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 145 ALEKGLPDPVLYAEKFTP 163
Db 145 ALEKGLPDPVLYAEKFTP 163

RESULT 6
Q8XN05 PRELIMINARY; PRT; 96 AA.
ID Q8XN05
AC Q8XN05;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein CPB0277.
GN CPB0277.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX MEDLINE=21664373; PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Onshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi A.,
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL; AP003186; BAB79983.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 96 AA; 10261 MW; 4FF6184BA168D076 CRC64;

Query Match 2.5%; Score 8; DB 16; Length 96;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 LAASFLLI 41
Db 9 LAASFLLI 16

RESULT 7
Q82CL7 PRELIMINARY; PRT; 120 AA.
ID Q82CL7
AC Q82CL7;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein.
GN SAV5328.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=33903;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NA-4680 / ATCC 31267 / NCTMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NA-4680 / ATCC 31267 / NCTMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.,
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL; AP005042; BAC73040.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 120 AA; 13549 MW; CB734655174F973 CRC64;

Query Match 2.5%; Score 8; DB 16; Length 120;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 263 LGGAVVSL 270
Db 34 LGGAVVSL 41

RESULT 8
Q8Z7B8 PRELIMINARY; PRT; 141 AA.
ID Q8Z7B8
AC Q8Z7B8;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical periplasmic protein STY1364.
GN STY1364 OR T1602.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltham J., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=2251367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kocoyianni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
DR EMBL; AL627270; CAD01632.1; -
DR EMBL; AB016839; AAO69231.1; -
KW Hypothetical protein; Complete proteome.

```

SQ SEQUENCE 141 AA; 15652 MW; BA6565B29A3B8844 CRC64;

Query Match 2.5%; Score 8; DB 16; Length 141;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 223 ALASISSY 230
Db 11 ALASISSY 18

RESULT 9

Q8N206 PRELIMINARY; PRT; 145 AA.

AC Q8N206; 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein FLJ90047.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Isogai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA Yamamoto Y., Wakematsu A., Nakamura Y., Kojima S., Nagahari K.,
RA Maehara Y., Ono T., Okano K., Yoshikawa Y., Aochioka S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Nishimura Y.,
RT "NEBO human cDNA sequencing project."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK074528; BAC11041.1; -
KW Hypothetical protein.
SQ SEQUENCE 145 AA; 16153 MW; EP4C8E658EAC2E CRC64;

Query Match 2.5%; Score 8; DB 4; Length 145;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 197 LSTPAPL 204
Db 42 LSTPAPL 49

RESULT 10

Q914L7 PRELIMINARY; PRT; 227 AA.

AC Q914L7; 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein PA118.
GN PA118.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Huftagie W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Labig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reiter J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
DR EMBL; AE004542; AAC04507.1; -

DR PIR: B83505; B83505.

KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 227 AA; 25612 MW; FAAD7BC0A50B9EC CRC64;

Query Match 2.5%; Score 8; DB 16; Length 227;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 98 RVGLVGL 105
Db 105 RVGLVGL 112

RESULT 11

Q54145 PRELIMINARY; PRT; 253 AA.

AC Q54145; 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Thioesterase.
OS Streptomyces fradiae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1906;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TS9235;
RX MEDLINE=95075319; PubMed=7984112;
RA Merson-Davies L.A., Cundliffe E.;
RT "Analysis of five tylosin biosynthetic genes from the tyl1BA region of
RT the Streptomyces fradiae genome."
RL Mol. Microbiol. 13:349-355(1994).
DR EMBL; U08223; AAA21345.1; -
DR PIR; S49055; S49055.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0016788; F:hydrolase activity, acting on ester bonds; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR InterPro; IPR000379; Ser_estcr.
DR InterPro; IPR001031; Thioesterase.
DR Pfam; PF00975; Thioesterase; 1.
SQ SEQUENCE 253 AA; 27684 MW; 8AB752B2CDB1450F CRC64;

Query Match 2.5%; Score 8; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 211 LTTGAPAL 218
Db 213 LTTGAPAL 220

RESULT 12

Q83P96 PRELIMINARY; PRT; 338 AA.

AC Q83P96; 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Periplasmic binding protein component of Pn transporter (Periplasmic
DE binding protein component of phosphonate transporter).
OS Shigella flexneri.
GN PHND OR SP4118 OR S3612.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12364590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,

RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
 RA Yu J.,
 RT "genome sequence of *Shigella flexneri* 2a: insights into pathogenicity
 RT through comparison with genomes of *Escherichia coli* K12 and O157.",
 RL Nucleic Acids Res. 30:4432-4441(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=2457T / ATCC 700930 / serotype 2a;
 RX MEDLINE=22590274; PubMed=12704152;
 RA Wei J., Goldberg M.B., Burtand V., Venkatesan M.M., Deng W.,
 RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
 RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.D., Zhou S.,
 RA Schwartz D.C., Blattner F.R.,
 RT "Complete genome sequence and comparative genomics of *Shigella*
 RT *flexneri* serotype 2a strain 2457T.",
 RL Infect. Immun. 71:2775-2786(2003).
 DR EMBL: AE016989; AAP16656.1; -
 DR EMBL: AE015420; AAN45543.1; -
 DR GO: GO:0030288; C:periplasmic space (sensu Gram-negative Bact. . .; IEA.
 DR GO: GO:0015604; P:phosphonate transporter activity; IEA.
 DR GO: GO:0015716; P:phosphonate transport; IEA.
 DR InterPro: IPR005770; Phosphonate bind.
 DR TrIRPAMS: TrIR01038; 3A0109803R.1.
 KM Complete proteome.
 SQ SEQUENCE 338 AA; 37382 MW; 9A9C3C3861CB594 CRC64;
 Query Match 2.5%; Score 8; DB 16; Length 338;
 Best Local Similarity 100.0%; Pred. No. 62;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 296 SPLIIGDP 303
 Db 225 SPLIIGDP 232

RESULT 13
 ID Q7YTN3 PRELIMINARY; PRT; 382 AA.
 AC Q7YTN3;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE C. elegans SRW-42 protein (Corresponding sequence F14F8.11).
 GN SRW-42.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodermata; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "genome sequence of the nematode *C. elegans*: A platform for
 RT investigating biology.",
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Lloyd C.R.,
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Z92782; CAE17813.1; -
 SQ SEQUENCE 382 AA; 43552 MW; 492BCAA4F37C4D1C CRC64;
 Query Match 2.5%; Score 8; DB 5; Length 382;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 56 RVLLSLFI 63
 Db 97 RVLLSLFI 104

RESULT 14
 ID Q9UR09 PRELIMINARY; PRT; 417 AA.
 AC Q9UR09;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Cell wall synthesis protein PSU1.
 GN PSU1 OR PSU1+.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetaceae; Schizosaccharomycetaceae;
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972H-;
 RA Cadden E., Lelaure V., Galibert F., McDougall R.C., Rajandream M.A.,
 RA Barrett B.G.,
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972H-;
 RX MEDLINE=9939312; PubMed=10462482;
 RA Omi K., Sonoda H., Nagata K., Sugita K.,
 RT "Cloning and characterization of psu1+, a new essential fission yeast
 RT gene involved in cell wall synthesis.",
 RL Biochem. Biophys. Res. Commun. 262:368-374(1999).
 DR EMBL: AL136078; CAB6613.1; -
 DR EMBL: AB009980; BAA83907.1; -
 DR PIR: J07092; J07092.
 DR InterPro: IPR005556; SUN.
 DR Pfam: PF03856; SUN; 1.
 SQ SEQUENCE 417 AA; 41750 MW; 72BCAE2226208CB CRC64;
 Query Match 2.5%; Score 8; DB 3; Length 417;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 208 IALLTTGA 215
 Db 9 IALLTTGA 16

RESULT 15
 ID Q9LAP6 PRELIMINARY; PRT; 468 AA.
 AC Q9LAP6;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE DNA mismatch repair protein (Fragment).
 GN MTHL.
 OS Staphylococcus aureus.
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=1280;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COL;
 RX MEDLINE=20031141; PubMed=10566865.
 RA de Lencastre H., Wu S.W., Pinho M.G., Ludovice A.M., Filipe S.,
 RA Gardete S., Sobral R., Gill S., Chung M., Tomasz A.,
 RT "Antibiotic resistance as a stress response: complete sequencing of a
 RT large number of chromosomal loci in *Staphylococcus aureus* strain COL
 RT that impact on the expression of resistance to methicillin.",
 RL Microb. Drug Resist. 5:163-175(1999).
 DR EMBL: AL131755; CAB82461.1; -
 DR HSSP: P23367; IBKN.
 DR GO: GO:0005524; F:ATP binding; IEA.
 DR GO: GO:0006298; P:mismatch repair; IEA.
 DR InterPro: IPR003594; ATPbind ATPase.
 DR InterPro: IPR002099; DNA_mis_repair.
 DR Pfam: PF01119; DNA_mis_repair; 1.

DR Pfam; PF02518; HATPase_C; 1.
DR SMART; SM00387; HATPase_C; 1.
DR TIGRFAMs; TIGR00585; multi; 1.
DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
FT NON_TER 468 468
SQ SEQUENCE 468 AA; 52987 MW; 87BADBD50109FBE CRC64;

Query Match 2.5%; Score 8; DB 2; Length 468;
Best Local Similarity 100.0%; Pred.No. 82;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 223 ALASISY 230
|||
Db 99 ALASISY 106

Search completed: February 23, 2004, 19:34:47
Job time : 47 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2004, 18:16:40 ; Search time 98 Seconds

(without alignments)
922.604 Million cell updates/sec

Title: US-09-936-456-2

Perfect score: 1666
Sequence: 1 MTLNMGVLPFPQPRHAGF.....GDPHKQALPDLKCTTTL 320

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	1666	100.0	320 3 AAB18992	Abd18992 Amino aci
2	1666	100.0	320 3 AAB26325	Abd26325 Human CAS
3	1660	99.6	320 5 AAB390126	Abd390126 Human pol
4	1660	99.6	370 7 ADD19303	Add19303
5	1253	75.2	242 3 AAB52095	Aab52095 Human sec
6	908	54.5	262 3 AAB26326	Aab26326 CASB618 P
7	883	53.0	343 5 ABB97510	Abb97510 Novel hum
8	874	52.5	343 4 AAB86071	Abd86071 Human INT
9	874	52.5	343 6 ABOU08370	Abou08370 Amino aci
10	874	52.5	343 6 ABOU08370	Abou08370 Amino aci
11	697.5	41.9	298 6 ADA55266	Ada55266 Human pro
12	460.5	27.6	445 4 ABOC3520	Aboc3520 Novel hum
13	458	27.5	112 4 AAB66073	Aab66073 Human INT
14	458	27.5	112 6 ABO32727	Abob32727 Secreted
15	276.5	16.6	383 4 AAB66060	Abab66060 COEEL.3 r
16	276.5	16.6	383 6 ABO32714	Abob32714 Secreted
17	254.5	15.3	363 4 AAB60704	Abb60704 Drosophila
18	254.5	15.3	363 4 AAB60704	Abb60704 Drosophila
19	254	15.2	103 7 ADD19262	Add19262 Human sec
20	242.5	14.6	253 3 AAB52094	Aab52094 Gene 44 h
21	110	6.6	450 6 AABU21931	Abu21931 Protein e
22	109	6.5	359 4 AAB20381	Abab20381 Soybean s
23	108	6.4	648 6 ABOU32147	Abu32147 Protein e
24	107	6.4	671 6 ABOU28385	Abu28385 Protein e
25	106.5	6.4	374 4 AAB20379	Abd20379 Corn ster

26	105.5	6.3	473 2 AAY38688	Aay38688 Neisseria
27	105.5	6.3	473 3 AAY75470	Aay75470 Neisseria
28	105.5	6.3	473 6 ABB7672	Abb7672 N. gonorr
29	104	6.2	364 5 AAG97436	Aag97436 S. garzin
30	104	6.2	652 4 AAB38223	Aab38223 Salmonella
31	104	6.2	652 6 ABOU47915	Abou47915 Protein e
32	103.5	6.2	428 6 ABOU21684	Abu21684 Protein e
33	103	6.2	98 4 AAG76780	Aag76780 Human col
34	101.5	6.1	428 6 AAB67285	Abm67285 Phototrab
35	101.5	6.1	465 6 ABOU48052	Abu48052 Protein e
36	100.5	6.0	431 4 AAB59558	Aab59558 Drosophila
37	100.5	6.0	585 2 AAY38798	Aay38798 Neisseria
38	99.5	6.0	477 4 AAB38247	Aab38247 Salmonella
39	99.5	6.0	533 6 ADA36605	Aaa36605 Acinetoba
40	99.5	6.0	558 2 AAY38799	Aay38799 Neisseria
41	99	5.9	612 6 ABOU31012	Abu31012 Protein e
42	98.5	5.9	317 3 AAY57097	Aay57097 UDP-glucu
43	98.5	5.9	464 4 AAB52659	Abd52659 Escherich
44	98.5	5.9	481 6 ABOU45403	Abu45403 Protein e
45	98	5.9	463 4 AAB34551	Aab34551 E. coli c

ALIGNMENTS

RESULT 1	
AAB18992	AAB18992 standard; protein; 320 AA.
AC	
XX	AAB18992;
DT	08-FEB-2001 (first entry)
XX	
DE	Amino acid sequence of a human transmembrane protein.
XX	
KM	Human; transmembrane protein; cell proliferation disorder; myeloma;
KM	reproductive disorder; smooth muscle disorder; neurological disorder;
KM	arteriosclerosis; leukemias; acquired immunodeficiency syndrome; AIDS;
KM	allergy; ovulatory defect; angina; hypertension; stroke; epilepsy;
KM	Alzheimer's disease; Tourette's disorder.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	Modified-site 84
FT	/note= "potential glycosylation site"
FT	Modified-site 86
FT	/note= "potential phosphorylation site"
FT	Modified-site 96
FT	/note= "potential phosphorylation site"
FT	Modified-site 109
FT	/note= "potential glycosylation site"
FT	Modified-site 121
FT	/note= "potential glycosylation site"
FT	Modified-site 131
FT	/note= "potential phosphorylation site"
FT	Modified-site 156
FT	/note= "potential glycosylation site"
FT	Modified-site 280
FT	/note= "potential phosphorylation site"
FT	Modified-site 285
FT	/note= "potential phosphorylation site"
PD	WO200056891-A2.
XX	
XX	28-SEP-2000.
XX	
XX	22-MAR-2000; 2000WO-US007817.
XX	
XX	22-MAR-1999; 99US-0125537P.
XX	16-JUN-1999; 99US-0139565P.
PA	(INCY-) INCYTE PHARM INC.

XX Yue H, Lai P, Tang YT, Hillman JL, Reddy R, Bandman O;
 PI Baughn MR, Lu DAM, Azimzai Y, Yang J;
 XX WPI: 2000-579485/54.
 DR N-PSDB; AAA96505.
 XX
 PT New human transmembrane proteins are used to treat a disease or condition
 associated with decreased expression of functional HTMP e.g. Tourette's
 disorder, angina and leukemia.
 XX
 PS Claim 1; Page 109; 130pp; English.
 CC The present sequence represents a human transmembrane proteins (HTMP).
 CC Agonists and antagonists of the protein are used to treat a disease or
 CC condition associated with overexpression of the protein. Diseases and
 CC conditions which can be treated include cell proliferation,
 CC immunological, reproductive, smooth muscle and neurological disorders
 CC e.g. arteriosclerosis, myeloma, leukaemia, acquired immunodeficiency
 CC syndrome (AIDS), allergies, ovulatory defects, angina, hypertension,
 CC stroke, Alzheimer's disease, epilepsy and Tourette's disorder. The
 CC polynucleotides may be used to detect and quantify gene expression in
 CC biopsied tissues where protein expression may be correlated with disease
 CC e.g. to determine absence, presence or excess expression of HTMP or to
 CC monitor regulation of HTMP expression during therapeutic intervention
 XX
 SQ Sequence 320 AA;
 Query Match 100.0%; Score 1666; DB 3; Length 320;
 Best Local Similarity 100.0%; Pred. No. 2.1e-173;
 Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTLNNGVLPFYPPQPRHAGFSVPLLIIVFLALAAFPLLIPGIRGSRMFWLVRLIS 60
 DB 1 MTLNNGVLPFYPPQPRHAGFSVPLLIIVFLALAAFPLLIPGIRGSRMFWLVRLIS 60
 QY 61 LFIGAETVAVHFSAEWVGVTNTNTSYKAFSAARVAVRGVLVGEINITLTGTVPVQL 120
 DB 61 LFIGAETVAVHFSAEWVGVTNTNTSYKAFSAARVAVRGVLVGEINITLTGTVPVQL 120
 QY 121 NETIDYNEQFTWRLKENYAAYANALEKGLPDPVLYIAEKFTPSSPCGLYHQYHLAGHYA 180
 DB 121 NETIDYNEQFTWRLKENYAAYANALEKGLPDPVLYIAEKFTPSSPCGLYHQYHLAGHYA 180
 QY 181 SATLWVAFCEWLLSNVLSTPAPLYGGLALTTGAPALFGVFAALASISSVPLCPRLGSS 240
 DB 181 SATLWVAFCEWLLSNVLSTPAPLYGGLALTTGAPALFGVFAALASISSVPLCPRLGSS 240
 QY 241 ALTYOYGAAPFWTLATGVLCTFLGAVVSIQYVRPSALRTLDDQAKDCSQERGSSPILL 300
 DB 241 ALTYOYGAAPFWTLATGVLCTFLGAVVSIQYVRPSALRTLDDQAKDCSQERGSSPILL 300
 QY 301 GDPPLHQAALPDLKCIITNL 320
 DB 301 GDPPLHQAALPDLKCIITNL 320
 RESULT 2
 AAB26325
 ID AAB26325 standard; protein; 320 AA.
 AC AAB26325;
 XX
 DT 11-JAN-2001 (first entry)
 XX
 DE Human CASB618 protein.
 XX
 KW Epitope; human; CASB618; chromosome 15; tumour; vaccine; cancer; ovarian;
 KW color; autoimmune disease; HLA_A0201.
 XX
 OS Homo sapiens.
 XX
 PN WO200053748-A2.

XX 14-SEP-2000.
 XX
 XX 09-MAR-2000; 2000WO-EP002048.
 XX
 XX 11-MAR-1999; 99GB-00005607.
 XX 01-SEP-1999; 99GB-00020590.
 XX
 PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX
 PI Bruck CEM, Cascart J, Coche T, Vinals Y De Bassolac.
 XX
 DR WPI: 2000-572268/53.
 DR N-PSDB; AAA94623.
 XX
 PT New human CASB618 polypeptide, useful as a vaccine for prophylactic and
 PT therapeutic treatment of cancers, particularly ovarian or colon cancer,
 PT autoimmune diseases and related conditions.
 XX
 PS Claim 1; Page 61-62; 76pp; English.
 CC The present sequence is human CASB618 protein. The gene for human CASB618
 CC is thought to be located on chromosome 15. The present protein and
 CC epitopes of this protein (see AAB26327 to AAB26399) are useful in
 CC diagnosing the occurrence of tumour cells and in vaccines for prophylactic
 CC and therapeutic treatment of cancers, particularly ovarian or colon
 CC cancer, autoimmune diseases and related conditions
 XX
 SQ Sequence 320 AA;
 Query Match 100.0%; Score 1666; DB 3; Length 320;
 Best Local Similarity 100.0%; Pred. No. 2.1e-173;
 Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTLNNGVLPFYPPQPRHAGFSVPLLIIVFLALAAFPLLIPGIRGSRMFWLVRLIS 60
 DB 1 MTLNNGVLPFYPPQPRHAGFSVPLLIIVFLALAAFPLLIPGIRGSRMFWLVRLIS 60
 QY 61 LFIGAETVAVHFSAEWVGVTNTNTSYKAFSAARVAVRGVLVGEINITLTGTVPVQL 120
 DB 61 LFIGAETVAVHFSAEWVGVTNTNTSYKAFSAARVAVRGVLVGEINITLTGTVPVQL 120
 QY 121 NETIDYNEQFTWRLKENYAAYANALEKGLPDPVLYIAEKFTPSSPCGLYHQYHLAGHYA 180
 DB 121 NETIDYNEQFTWRLKENYAAYANALEKGLPDPVLYIAEKFTPSSPCGLYHQYHLAGHYA 180
 QY 181 SATLWVAFCEWLLSNVLSTPAPLYGGLALTTGAPALFGVFAALASISSVPLCPRLGSS 240
 DB 181 SATLWVAFCEWLLSNVLSTPAPLYGGLALTTGAPALFGVFAALASISSVPLCPRLGSS 240
 QY 241 ALTYOYGAAPFWTLATGVLCTFLGAVVSIQYVRPSALRTLDDQAKDCSQERGSSPILL 300
 DB 241 ALTYOYGAAPFWTLATGVLCTFLGAVVSIQYVRPSALRTLDDQAKDCSQERGSSPILL 300
 QY 301 GDPPLHQAALPDLKCIITNL 320
 DB 301 GDPPLHQAALPDLKCIITNL 320
 RESULT 3
 ABB90126
 ID ABB90126 standard; protein; 320 AA.
 AC ABB90126;
 XX
 DT 24-MAY-2002 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 2502.
 XX
 KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KW antiallergic; hepatocytic; antidiabetic; antiinflammatory; anticancer;
 KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;

KM neurological disease; infection; human; secreted protein.
 XX Homo sapiens.
 OS
 XX
 PN WO200190304-A2.
 XX
 XX 29-NOV-2001.
 PD
 XX 18-MAY-2001; 2001WO-US016450.
 PF
 XX 19-MAY-2000; 2000US-020551SP.
 PR
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Blaise CE, Rosen CA;
 PI
 XX WPI: 2002-122018/16.
 DR N-PSDB; ABL90535.
 DR
 XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
 PT prevention of neural, immune system, muscular, reproductive,
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
 PT disorders.
 PS
 XX Claim 11; SEQ ID NO 2502; 2081bp + Sequence Listing; English.
 PS
 XX The invention relates to novel genes (ABJ89449-ABJ90853) and proteins
 CC (ABJ89040-ABJ90444) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
 CC ovarian cancer, and other cancers of the adrenal gland, bone, bone marrow,
 CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, diabetes mellitus, Crohn's disease,
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
 CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing
 CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
 CC infectious diseases such as viral, bacterial, fungal and parasitic
 CC infections. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 CC
 XX
 SQ Sequence 320 AA;
 Query Match 99.6%; Score 1660; DB 5; Length 320;
 Best Local Similarity 99.7%; Pred. No. 9,6e-173;
 Matches 319; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 4
 ID ADD19303 standard, protein; 370 AA.
 AC
 XX ADD19303;
 AC
 XX 15-JUN-2004 (first entry)
 DE
 XX Human secreted protein from gene 4 #2.
 DE
 XX human secreted protein; cytosolic; antibacterial; virucide;
 KM neuroprotective; gynaecological; gastrointestinal; cardiac;
 KM cardiovascular; nephrotropic; antiinflammatory; muscular; Gen;
 KM respiratory; Gen; immunosuppressive; cerebroprotective; vasotropic;
 KM neutropic; antiallergic; cancer; bacterial infection; viral infection;
 KM neural disorder; immune system disorder; blood disorder;
 KM muscular disorder; reproductive disorder; gastrointestinal disorder;
 KM pulmonary disorder; cardiovascular disorder; renal disorder;
 KM inflammatory disorder; proliferative disorder; human.
 OS
 XX Homo sapiens.
 PS
 XX WO2003052377-A2.
 PN
 XX 26-JUN-2003.
 PD
 XX 06-NOV-2002; 2002WO-US035606.
 PF
 XX 07-NOV-2001; 2001US-0331046P.
 PR
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Rosen CA, Ruben SM;
 PI
 XX WPI: 2003-533050/50.
 DR N-PSDB; ADD19228.
 DR
 XX New isolated nucleic acids encoding signal transduction pathway component
 PT polypeptides, useful for diagnosing, treating, and/or preventing
 PT disorders, such as cancer, infections, cardiovascular and inflammatory
 PT diseases.
 PS
 XX Claim 11; SEQ ID NO 130; 554bp; English.
 PS
 XX The invention relates to an isolated nucleic acid molecule (cDNA)
 CC encoding a human secreted protein, representing one of 85 novel genes.
 CC Also included are recombinant vectors, host cells (expressing the
 CC protein), the secreted proteins (including their fragments, epitopes and
 CC homologues), an isolated antibody that binds specifically to the protein,
 CC diagnosing a pathological condition or susceptibility to a pathological
 CC condition (comprising determining the presence or absence of a mutation
 CC in the nucleic acid and diagnosing a condition based on the presence or
 CC absence of the mutation), diagnosing a pathological condition or
 CC susceptibility to a pathological condition (comprising determining the
 CC presence or amount of expression of the protein in a biological sample
 CC and diagnosing a condition based on the presence or amount of expression
 CC of the protein), preventing, treating or ameliorating a medical condition
 CC by administering the nucleic acid or protein to a mammalian subject,
 CC identifying a binding partner to the protein, the gene corresponding to
 CC the cDNA sequence, and identifying an activity in a biological assay
 CC (comprising expressing the nucleic acid in a cell, isolating the
 CC supernatant, detecting an activity in a biological assay and identifying
 CC the protein in the supernatant having the activity). The nucleic acids
 CC and proteins display the following activities: Cytostatic, antibacterial,
 CC virucide, Neuroprotective, Gynaecological, Gastrointestinal, Gen,
 CC Cardiac, Cardiovascular, Gen, Nephrotropic, Antiinflammatory, Muscular-
 CC Gen, Respiratory, Gen, Immunosuppressive, Cerebroprotective, Vasotropic,
 CC Neutropic, Antiallergic. The methods and compositions of the present
 CC invention are useful for diagnosing, treating, preventing and/or
 CC prognosticating disorders related to the novel polypeptides, such as
 CC cancer, bacterial or viral infections, and neural, immune system, blood,

CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
 CC renal, inflammatory or proliferative disorders (many examples of these
 CC diseases and disorders are given in the specification). The present
 CC sequence represents a novel secreted protein of the invention.

XX SQ Sequence 370 AA;

Query Match 99.6%; Score 1660; DB 7; Length 370;
 Best Local Similarity 99.7%; Pred. No. 1.2e-172;
 Matches 319; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MLLMNGVLEFYPQPHAGFSPVLLIVLPLALAASTLLPGLRGHSRWFMLVRYLLS 60
 DB 51 MLLMNGVLEFYPQPHAGFSPVLLIVLPLALAASTLLPGLRGHSRWFMLVRYLLS 110
 QY 61 LFIGAETVAHFSAEFVGTAVNTSYKAFSAARVTAAGLVLEGINTLTGTPVHQL 120
 DB 111 LFIGAETVAHFSAEFVGTAVNTSYKAFSAARVTAAGLVLEGINTLTGTPVHQL 170
 QY 121 NETIDYNEOPTWRLKENYAAEYANALEKGLDPVLYLAKEPTSSPGCLYHQYHLA 180
 DB 171 NETIDYNEOPTWRLKENYAAEYANALEKGLDPVLYLAKEPTSSPGCLYHQYHLA 230
 QY 181 SATLWVAFCEFWLLSNVLTSTPAPLYGGLALLTTGAFALFGYFALASISVPLCPLRGSS 240
 DB 231 SATLWVAFCEFWLLSNVLTSTPAPLYGGLALLTTGAFALFGYFALASISVPLCPLRGSS 290
 QY 241 ALTYGGAFFWTLATGVCLFLGAVVSLQYVPSALRTLLDQSAKDCQERGSPILL 300
 DB 291 ALTYGGAFFWTLATGVCLFLGAVVSLQYVPSALRTLLDQSAKDCQERGSPILL 350
 QY 301 GDPKHKQALPDLKCIITNL 320
 DB 351 GDPKHKQALPDLKCIITNL 370

RESULT 5
 AAB52095
 ID AAB52095 standard; protein; 242 AA.

XX AC AAB52095;

XX DT 21-FEB-2001 (first entry)

XX DE Human secreted protein sequence encoded by gene 44 SEQ ID NO:144.

XX KW Human; secreted protein; cytosolic; immunosuppressive; nocotropic;
 KW neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic;
 KW antiinflammatory; antiviral; anticonvulsant; antibacterial;
 KW antifungal; antiparasitic; cardiac; cancer; immune disease; allergy;
 KW cardiovascular disorder; wound healing; infection; neurological disease.

XX OS Homo sapiens.

XX PN MO200061596-A1.

XX PD 19-OCT-2000.

XX PF 06-APR-2000; 2000WO-US008983.

XX PR 09-APR-1999; 99US-0128703P.

XX PR 14-JAN-2000; 2000US-0176068P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PA (ROSE/) ROSEN C A.

XX PI Ruben SM, Komatsoulis G;

XX DR MPI: 2000-611865/58.

XX DR N-PSDB; AAC95564.

XX PT Fifty nucleic acid molecules encoding human secreted proteins, useful in
 XX the prevention, treatment and diagnosis of cancer, immune disorders,

PT cardiovascular disorders and neurological diseases.

XX Disclosure: Page 75-76; 505pp; English.

XX CC Polynucleotide sequences AAC95521 - AAC95570 represent cDNA encoding 50
 PS CC human secreted proteins AAB52012 - AAB52161. Sequences AAB52062 -
 XX CC AAB52103 represent alternative polypeptides encoded by the genes, and
 CC amino acid sequences with which they share homology. The genes and
 CC proteins have activities dependent on the tissues and cells in which they
 CC are expressed. Examples of their activities include cytosolic;
 CC immunosuppressive; nocotropic; neuroprotective; antiviral; antiallergic;
 CC hepatotropic; antidiabetic; antiinflammatory; antiviral; vulnary;
 CC anticonvulsant; antibacterial; antifungal; antiparasitic; and cardiac.
 CC The secreted proteins, polypeptides, antagonists and agonists may be
 CC useful in treating, preventing and/or diagnosing diseases and disorders
 CC such as cancer, particularly breast and ovarian cancer, and other cancers
 CC of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract,
 CC liver, lung, or urogenital. Immune disorders such as Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; cardiovascular disorders such as
 CC myocardial ischaemia; wound healing; neurological diseases such as
 CC cerebral anoxia and epilepsy; and infectious diseases such as viral,
 CC bacterial, fungal and parasitic infections may also be treated using the
 CC proteins and polynucleotides of the invention. Sequences AAC95512 -
 CC AAC95520 and AAB52011 are used in the isolation and characterisation of
 CC the proteins and polynucleotides of the invention

XX SQ Sequence 242 AA;

Query Match 75.2%; Score 1253; DB 3; Length 242;
 Best Local Similarity 99.6%; Pred. No. 2.4e-128;
 Matches 241; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 44 GIRGSRWFMLVRYLLSLEFAGETVAHFSAEFVGTAVNTSYKAFSAARVTAAGLV 103
 DB 1 GIRGSRWFMLVRYLLSLEFAGETVAHFSAEFVGTAVNTSYKAFSAARVTAAGLV 60
 QY 104 GLEGINTLTGTPVHQLNETIDYNEOPTWRLKENYAAEYANALEKGLDPVLYLAKEPTP 163
 DB 61 GLEGINTLTGTPVHQLNETIDYNEOPTWRLKENYAAEYANALEKGLDPVLYLAKEPTP 120
 QY 164 SSPGCLYHQYHLAAGVATLWVAFCEFWLLSNVLTSTPAPLYGGLALLTTGAFALGV 223
 DB 121 SSPGCLYHQYHLAAGVATLWVAFCEFWLLSNVLTSTPAPLYGGLALLTTGAFALGV 180
 QY 224 LASISSVPLCPLRGSSALTTGGAFFWTLATGVCLFLGAVVSLQYVPSALRTLLD 283
 DB 181 LASISSVPLCPLRGSSALTTGGAFFWTLATGVCLFLGAVVSLQYVPSALRTLLD 240
 QY 284 QS 285
 DB 241 QS 242

RESULT 6
 AAB26326
 ID AAB26326 standard; protein; 262 AA.

XX AC AAB26326;

XX DT 12-SEP-2003 (revised)

XX DT 11-JAN-2001 (first entry)

XX DE CASB618 protein.

XX KW Epitope; human; CASB618; chromosome 15; tumour; vaccine; cancer; ovarian;
 XX colon; autoimmune disease; HLA_A0201.

XX OS Homo sapiens.

XX OS Influenza virus.

XX OS Chimeric.

```

PN W0200053748-A2.
XX
XX 14-SEP-2000.
PD
XX
XX 09-MAR-2000; 2000MO-EP002048.
PE
XX
XX 11-MAR-1999; 99GB-00005607.
PR 01-SEP-1999; 99GB-00020590.
XX
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
PA
XX
XX Bruck CEM, Cascart J, Coche T, Vinals Y De Bassolac;
PI WPI: 2000-572268/53.
XX
XX New human CASB618 polypeptide, useful as a vaccine for prophylactic and
PT therapeutic treatment of cancers, particularly ovarian or colon cancer,
PT autoimmune diseases and related conditions.
XX
XX Example 7; Page 63; 76pp; English.
PS
XX The present invention relates to epitopes of human CASB618 protein (see
CC AAB26327 to AAB26399). The epitopes of AAB26327 to AAB26399 are useful in
CC diagnosing the occurrence of tumour cells and in vaccines for prophylactic
CC and therapeutic treatment of cancers, particularly ovarian or colon
CC cancer, autoimmune diseases and related conditions. The present sequence
CC is a protein derived from a chimeric gene constructed from the CASB618
CC gene carrying deletions of the N-terminus and C-terminus, with the
CC addition of an IFP (NS1 DNA sequence encoding the N-terminal 1-81 amino
CC acids of the NS1 protein of influenza virus) at the N-terminus, and a C-
CC terminal histidine tail. The gene encoding the present protein was cloned
CC and used to evaluate the effectiveness of tumour-specific antigens.
CC (Updated on 12-SEP-2003 to standardise OS field)
CC
XX
SQ Sequence 262 AA;
Query Match 54.5%; Score 908; DB 3; Length 262;
Best Local Similarity 97.2%; Pred. No. 1.7e-90;
Matches 173; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 69 AVHFSAEWVGVGNTNTSYKAFSAARVAVGLVLEGINTLTGTPVHQLNETIDYNE 128
DB 76 ALKRTWMEFVGNTNTSYKAFSAARVAVGLVLEGINTLTGTPVHQLNETIDYNE 135
QY 129 QFTWRLKENYAAEYANALEKGLPDPVLYIAEKFTSPSCGLYHOYHLAGHYASATLWYAF 188
DB 136 QFTWRLKENYAAEYANALEKGLPDPVLYIAEKFTSPSCGLYHOYHLAGHYASATLWYAF 195
QY 189 CFWLLSNVLLSTPAPLYGGIALITGAFALFGVFAALASISVPLCPRLGSSALTTOY 246
DB 196 CFWLLSNVLLSTPAPLYGGIALITGAFALFGVFAALASISVPLCPRLGSSALTTOY 253

```

RESULT 7
AAB97510
ID AAB97510 standard; protein; 343 AA.
XX
AC AAB97510;
XX
XX 27-JUN-2002 (first entry)
XX
XX Novel human protein SEQ ID NO: 778.
XX
XX Human; anti-naemic; viroterary; anti-inflammatory; immunomodulator;
XX anti-fertility; cerebroprotective; cytoskeletal; rheumatic; gene therapy;
XX neuroprotective; antiparkinsonian; protein therapy; EST;
XX expressed sequence tag.
XX
XX Homo sapiens.
XX
XX W0200222660-A2.
XX
XX 21-MAR-2002.

```

XX
XX 10-SEP-2001; 2001WO-US026015.
PF
XX
XX 11-SEP-2000; 2000US-00659671.
PR
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Tang YT, Liu C, Zhou F, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Weinman T, Drmanac RT;
XX
XX WPI: 2002-292408/33.
DR N-PSDB; AAM32696.
XX
XX An isolated polynucleotide for treating diseases associated with its
PT encoded polypeptide such as cancer and multiple sclerosis.
PT
XX
XX Claim 20; SEQ ID NO 778; 509pp; English.
PS
XX
XX The present invention provides the protein and coding sequences of 444
CC novel human proteins. These were isolated from expressed sequences tags
CC (ESTs). They can be used to stimulate cell growth, to regulate
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC e.g. in burn treatment, to regulate the immune system e.g. to treat
CC multiple sclerosis, to regulate the immune system e.g. to treat
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.
CC rheumatoid arthritis, and to treat nervous system disorders e.g.
CC Parkinson's disease. The present sequence is a protein of the invention
XX
XX
SQ Sequence 343 AA;
Query Match 53.0%; Score 883; DB 5; Length 343;
Best Local Similarity 58.8%; Pred. No. 1.3e-87;
Matches 173; Conservative 40; Mismatches 73; Indels 8; Gaps 3;
QY 9 PFYQPRHAGFVSPLLIVLVFLALASFLILPGIGHGRWFVYVLSLFGAIEIV 68
DB 9 PFYQAGKPFPPMDTLTASIMFLTALNTFVILPGIGHGRKRLFMRLRVVSLTIGAIL 68
QY 69 AVHFSAEWVGVGNTNTSYKAFSAARVAVGLVLEGINTLTGTPVHQLNETIDYNE 128
DB 69 AVHFSAEWVGVGNTNTSYKAFSAARVAVGLVLEGINTLTGTPVHQLNETIDYNE 128
QY 129 QFTWRLKENYAAEYANALEKGLPDPVLYIAEKFTSPSCGLYHOYHLAGHYASATLWYAF 188
DB 129 EFTWRLGENYAAEYAKALEKGLPDPVLYIAEKFTSPSCGLYHOYHLAGHYASATLWYAF 188
QY 189 CFWLLSNVLLSTPAPLYGGIALITGAFALFGV--FALASISVPLCPRLGSSALTTOY 246
DB 189 LCVLLANVWLSMPALVGGYMLATGIFQLATLFFSWATSLTSP-CPHLGASVLAETHH 247
QY 247 GAFFWTATGVLCFLIGAVVLSLOYVPSALRTLLDOSAK-----DCSQRGG 295
DB 248 GPALFWITLTGTLGVLLGLANAVAHMPQHRKAFENQSVDEPMLZMSPEEG 301

```

RESULT 8
AAB66071
ID AAB66071 standard; protein; 343 AA.
XX
AC AAB66071;
XX
XX 30-MAR-2001 (first entry)
XX
XX Human INTERCEPT 296.
XX
XX TANGO protein; INTERCEPT protein; neurological disorder;
XX central nervous system; focal brain disorder; bipolar affective disorder;
XX global diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;
XX senile dementia; Huntington's disease; amyotrophic lateral sclerosis;
XX Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;
XX neuropsychiatric; psychoactive substance use; anxiety.
XX

OS Homo sapiens.
 XX MO200077239-A2.
 XX 21-DEC-2000.
 PD 24-MAY-2000; 2000MO-US014858.
 XX 14-JUN-1999; 99US-00333159.
 XX (MILL-) MILLENNIUM PHARM INC.
 PA McCarthy SA, Fraser CC, Sharp JD, Barnes TM,
 XX WPI; 2001-032313/04.
 DR N-PSDB; AAF45133, AAF45134.
 XX
 PT TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for
 PT screening assays and diagnostic assays and for the treatment of
 PT neurological diseases such as Alzheimer's, Parkinson's and Huntington's
 PT disease.
 PS Claim 8; Fig 7; 359pp; English.
 XX
 CC The present invention relates to TANGO or INTERCEPT proteins and coding
 CC sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66011-B66057,
 CC AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding
 CC sequences are useful for the treatment of neurological disorders such as
 CC central nervous system (CNS) disorders, CNS-related disorders, focal
 CC brain disorders, global diffuse cerebral disorders and other neurological
 CC and cerebrovascular disorders. The CNS disorders include Alzheimer's
 CC disease, senile dementia, Huntington's disease, amyotrophic lateral
 CC sclerosis, Parkinson's, Gilles de la Tourette's syndrome, autonomic
 CC function disorders such as hypertension and sleep disorders,
 CC neuropsychiatric disorders, psychoactive substance use disorders,
 CC anxiety, and bipolar affective disorder
 CC
 SQ Sequence 343 AA;
 Query Match 52.5%; Score 874; DB 4; Length 343;
 Best Local Similarity 58.5%; Pred. No. 1.3e-86;
 Matches 172; Conservative 40; Mismatches 74; Indels 8; Gaps 3;
 QY 9 PFYQPPHAGFSVPLIVLIVFLALASFLILPGIRGSRFWLVRLLSFLGAEIV 68
 DB 9 PFYAPKPTPFMDTLASIMIFLALATFIVILPGIRKRLFWLKVTSFLFGAIL 68
 QY 69 AAFSAEWFVGTANTNTSYKAPSAARVATRVGLVGLSGINITLTGTPVHQLNETIDYNE 128
 DB 69 AAFSEMSVGVGSNTSYKAFSSSEWISADIGLVGCVNITLTGTPVQQLNETINYE 128
 QY 129 CFTWRLEKYAAEYANALEKGLPDPVLYIAEKFTPPSPCGLYHQYHLAGYASATLVMAF 188
 DB 129 EFTWRLEKYAAEYAECKALEKGLPDPVLYIAEKFTPPSPCGLYRKYRAGHTAMLVMAF 188
 QY 189 CFWLSINVLSTPADLYGGLALITGGAFAFGV--PALASTSVPLCPRLGSSALTTQY 246
 DB 189 LCMILANVLSMPVLYVGGYMLLATGIFQLATLFFSATSTLSP-CPILHGLASVLTHTH 247
 QY 247 GAFFWTLATGVLCLFLGAVVSLQYVRPSALRTLLQOSAK-----DQSGERG 295
 DB 247 GAFFWTLTTLTGILCVLLGLMAVAHAKMPHRLKAFPNOSVDEDEMLWSPDEBG 301
 RESULT 9
 ABU08370 ID ABU08370 standard; protein; 343 AA.
 AC ABU08370;
 XX 03-JUN-2003 (first entry)
 DT
 XX Amino acid sequence for human INTERCEPT 296.
 DE

XX Human; INTERCEPT 296; TANGO; cellular process; tissue typing;
 KW forensic biology; cellular protease activity; cell interaction;
 KW development; blood disorder; hematopoietic cell-related disorder;
 KW growth; cell proliferation; cell differentiation; gamma delta T-cell;
 KW immune system; metabolic disorder; homeostatic disorder; anaemia;
 KW developmental bone disorder; osteoporosis; bacterial infection;
 KW thrombocytopenia; renal failure; liver disease; cystic fibrosis;
 KW obesity; cancer; bronchitis; asthma; emphysema; pulmonary oedema;
 KW respiratory distress syndrome; osteopathic; antibacterial; antineoplastic;
 KW thrombolytic; nephrotoxic; antioesity; hepatotropic; cytoskeletal;
 KW antiinflammatory; antiasthmatic.
 XX
 OS Homo sapiens.
 XX US2002182675-A1.
 XX
 XX 05-DEC-2002.
 PD 25-OCT-2001; 2001US-00042431.
 XX 14-JUN-1999; 99US-00333159.
 XX 24-MAY-2000; 2000US-00578063.
 XX
 XX (MILL-) MILLENNIUM PHARM INC.
 PA McCarthy SA, Fraser CC, Sharp JD, Barnes TM,
 XX WPI; 2003-328617/31.
 DR N-PSDB; ABX94107.
 XX
 PT New TANGO and INTERCEPT proteins, useful as modulating agents in
 PT regulating a variety of cellular processes, in chromosome mapping, in
 PT tissue typing, and in forensic biology.
 XX
 PS Disclosure; Fig 7A-7C; 232pp; English.
 CC
 CC The present invention relates to the isolation of novel TANGO or
 CC INTERCEPT proteins, and the polynucleotide sequences encoding them. The
 CC invention discloses sequences for human TANGO 202, TANGO 234, TANGO 265,
 CC TANGO 273, TANGO 286, TANGO 294 and INTERCEPT 296. The sequences for
 CC murine TANGO 202 and TANGO 273 are also provided. The TANGO polypeptides
 CC and the polynucleotide sequences encoding them are useful as modulating
 CC agents in regulating a variety of cellular processes. The polynucleotide
 CC sequences are useful as primers or hybridisation probes for the detection
 CC of nucleic acids encoding TANGO polypeptides; chromosome mapping; tissue
 CC typing, and in forensic biology. TANGO 202 can be used to diagnose,
 CC prevent or treat disorders relating to aberrant cellular protease
 CC activity, inappropriate interaction of cells with mediators,
 CC inappropriate development, and blood and hematopoietic cell-related
 CC disorders. TANGO 234 can be used to modulate growth, proliferation,
 CC survival, differentiation, and activity of gamma delta T-cells. TANGO 265
 CC can be used to prevent, diagnose and treat disorders characterised by
 CC aberrant organisation or development of a tissue or organ, and for
 CC modulating differentiation of cells of the immune system. TANGO 273 is
 CC useful for diagnosing, treating or preventing e.g. metabolic, homeostatic
 CC and developmental bone disorders (e.g. osteoporosis), and bacterial
 CC infection. TANGO 286 is useful for treating or preventing e.g. anaemia,
 CC thrombocytopenia, renal failure or liver disease. TANGO 294 is useful
 CC for treating or preventing e.g. cystic fibrosis or obesity. INTERCEPT 296
 CC is useful for diagnosing, treating or preventing e.g. cancer,
 CC bronchitis, cystic fibrosis, asthma, emphysema, pulmonary oedema, or
 CC adult and infant respiratory distress syndromes. The present sequence
 CC represents human INTERCEPT 296
 CC
 SQ Sequence 343 AA;
 Query Match 52.5%; Score 874; DB 6; Length 343;
 Best Local Similarity 58.5%; Pred. No. 1.3e-86;
 Matches 172; Conservative 40; Mismatches 74; Indels 8; Gaps 3;
 QY 9 PFYQPPHAGFSVPLIVLIVFLALASFLILPGIRGSRFWLVRLLSFLGAEIV 68
 DB 9 PFYAPKPTPFMDTLASIMIFLALATFIVILPGIRKRLFWLKVTSFLFGAIL 68

```

Db      9 PFYAGPKPTFPMDTLASIMIFLALATFIVILPGIRKTRLFMLLRVTSLFGAIL 68
Qy      69 AVHPSAEWVGTVNNTSYKAFSAARVTARVGLVGLGINTLTGTPVQNLNTIDYNE 128
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      69 AVNFSSEWVGQVSTNTSYKAFSSEWISADIGLQVGLGVNLTITGTVQQLNETINNE 128
Qy      129 QFTWRLKENYAAEVANALEKGLPDPVLYLAKEFTPSSPCGLYHQYHLAGHYASATLWVAF 188
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      129 EFTWRLGENYAAEBCAKALEKGLPDPVLYLAKEFTPSPCGLYRQYRLAGHYTSAMLWVAF 188
Qy      189 CFMLLSNVLSTPAPLYGGLALLTGAFALFGV--FALASISSVPLCGPLRGSSALTTQY 246
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      189 LCMLLANVWLSNPVLYVGGYMLLATGIFQLALLPFSWATSLTSP-CPHLGASVLATHH 247
Qy      247 GAAPFWTLATGVLCFLGAVVSLQYVRPSALRTLLDOSAK-----DCSQERGG 295
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      248 GFAPWTLTTLTGCLLVGLAMAVAHRMQPHRLKAFNQSVDDEPMLMSPEEG 301

RESULT 10
AB032725
ID      AB032725 standard; protein; 343 AA.
XX      AC      AB032725;
XX      DT      17-SEP-2003 (first entry)
XX      DE      Secreted polypeptide-related protein #100.
XX      KW      Human; TANGO; INTERCEPT; secreted polypeptide; immune disorder;
        KW      hormonal disorder; proliferative disorder; cancer; thyroid disorder;
        KW      diabetes; multiple sclerosis; lupus; neurological disorder; anemia;
        KW      Alzheimer's disease; Parkinson's disease; cardiovascular disorder;
        KW      myocardial infarction; congestive heart disease; blood platelet disorder;
        KW      thrombocytopenia; blood vessel; atherosclerosis; vasculitis.
XX      OS      Homo sapiens.
XX      PN      US2003022279-A1.
XX      PD      30-JUN-2003.
XX      PF      12-JAN-2001; 2001US-00759130.
XX      PR      14-JUN-1999; 99US-00333159.
        PR      29-JUN-1999; 99US-00342364.
        PR      10-SEP-1999; 99US-00393996.
        PR      19-OCT-1999; 99US-00420707.
        PR      07-JAN-2000; 2000US-00479249.
        PR      27-APR-2000; 2000US-00559497.
        PR      24-MAY-2000; 2000US-00578063.
        PR      16-JUN-2000; 2000US-00596194.
        PR      23-JUN-2000; 2000US-00602871.
        PR      30-JUN-2000; 2000US-00608452.
XX      PA      (FRAS/) FRASER C C.
        PA      (BARN/) BARNES T M.
        PA      (SHAR/) SHARP J D.
        PA      (KIRS/) KIRST S J.
        PA      (MYER/) MYERS P S.
        PA      (LEIB/) LEIBY K R.
        PA      (HOLT/) HOLTZMAN D A.
        PA      (MCCA/) MCCARTHY S A.
        PA      (WRIE/) WRIGHTON N.
        PA      (MACK/) MACKAY C R.
        PA      (GOOD/) GOODEARL A D J.
XX      PI      Fraser CC, Barnes TM, Sharp JD, Kirst SJ, Myers PS, Leiby KR;
        PI      Holtzman DA, McCarthy SA, Wrigton N, Mackay CR, Goodearl AD;
XX      DR      WPI; 2003-456290/43.
XX      DR      N-PSDB; ACD66786, ACD66787.
XX

```

```

PT      New nucleic acid molecule encoding a secreted protein (e.g. TANGO 202,
PT      TANGO 210 or INTERCEPT 217), useful for diagnosing, preventing or
PT      treating disorders such as cancer, diabetes or atherosclerosis, and in
PT      forensic biology.
PS      Claim 9; Fig 30A-30C; 462bp; English.
XX
XX      The invention relates to secreted polypeptide-related proteins and
CC      nucleic acids (TANGO and INTERCEPT proteins and nucleic acids). The
CC      nucleic acids, proteins and antibodies specific to the proteins are
CC      useful in screening assays, predictive medicine (e.g. diagnostic assays,
CC      prognostic assays, monitoring clinical trials and pharmacogenetics) and
CC      prophylactic and therapeutic methods. The sequences are used in
CC      diagnosing, preventing or treating proliferative disorders (e.g.
CC      cancers), hormonal disorders (e.g. diabetes or thyroid disorders), immune
CC      disorders (e.g. multiple sclerosis or lupus), neurological disorders
CC      (e.g. Alzheimer's disease or Parkinson's disease), cardiovascular
CC      disorders (e.g. myocardial infarction or congestive heart disease), blood
CC      platelet disorders (e.g. thrombocytopenia or anaemia) and disorders
CC      involving blood vessels (e.g. atherosclerosis or vasculitis). The nucleic
CC      acids may also be used in chromosome mapping, tissue typing and forensic
CC      biology, and as surrogate markers. This sequence represents a secreted
CC      polypeptide-related protein of the invention. Note: The sequence data for
CC      this patent was obtained in electronic format directly from USPTO at
CC      seqdata.uspto.gov/sequence.html
XX
XX      Sequence 343 AA;
SQ
XX      Query Match          52.5%; Score 874; DB 6; Length 343;
XX      Best local similarity 58.5%; Pred. No. 1.3e-86;
XX      Matches 172; Conservative 40; Mismatches 74; Indels 8; Gaps 3;
XX
Qy      9 PFYAPKPTFPMDTLASIMIFLALATFIVILPGIRKTRLFMLLRVTSLFGAIL 68
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      9 PFYAPKPTFPMDTLASIMIFLALATFIVILPGIRKTRLFMLLRVTSLFGAIL 68
Qy      69 AVHPSAEWVGTVNNTSYKAFSAARVTARVGLVGLGINTLTGTPVQNLNTIDYNE 128
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      69 AVNFSSEWVGQVSTNTSYKAFSSEWISADIGLQVGLGVNLTITGTVQQLNETINNE 128
Qy      129 QFTWRLKENYAAEVANALEKGLPDPVLYLAKEFTPSSPCGLYHQYHLAGHYASATLWVAF 188
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      129 EFTWRLGENYAAEBCAKALEKGLPDPVLYLAKEFTPSPCGLYRQYRLAGHYTSAMLWVAF 188
Qy      189 CFMLLSNVLSTPAPLYGGLALLTGAFALFGV--FALASISSVPLCGPLRGSSALTTQY 246
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      189 LCMLLANVWLSNPVLYVGGYMLLATGIFQLALLPFSWATSLTSP-CPHLGASVLATHH 247
Qy      247 GAAPFWTLATGVLCFLGAVVSLQYVRPSALRTLLDOSAK-----DCSQERGG 295
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      248 GFAPWTLTTLTGCLLVGLAMAVAHRMQPHRLKAFNQSVDDEPMLMSPEEG 301

RESULT 11
ADA55266
ID      ADA55266 standard; protein; 298 AA.
XX      AC      ADA55266;
XX      DT      20-NOV-2003 (first entry)
XX      DE      Human protein. SEQ ID 2834.
XX      KW      Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
        KW      Gene Therapy; human; secretory protein; membrane proteins; cancer;
        KW      inflammatory disease; osteoporosis; neurological disease.
XX      OS      Homo sapiens.
XX      PN      EP1293569-A2.
XX      PD      19-MAR-2003.
XX

```

PF 21-MAR-2002; 2002EP-00006586.
 XX
 PR 14-SEP-2001; 2001JP-00328381.
 PR 24-JAN-2002; 2002US-0350435P.
 XX
 XX (HELI-) HELIX RES INST.
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX
 XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;
 XX
 DR WPI: 2003-395539/38.
 DR N-PSDB; ADA53627.
 XX
 PT New polynucleotides encoding full-length polypeptides, e.g. secretory
 PT and/or membrane proteins, useful for developing medicines for diseases in
 PT which the gene is involved, or as target molecules for gene therapy.
 XX
 PS Claim 14; SEQ ID NO 2834; 205bp; English.
 XX
 CC The present invention relates to novel human secretory or membrane
 CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
 CC ADA54071). The coding sequences are useful in the gene therapy of
 CC diseases caused by abnormalities of the proteins, e.g. cancer,
 CC inflammatory diseases, osteoporosis or neurological disease.
 XX
 SQ Sequence 298 AA;
 Query Match 41.9%; Score 697.5; DB 6; Length 298;
 Best Local Similarity 48.6%; Pred. No. 2.3e-67;
 Matches 143; Conservative 32; Mismatches 66; Indels 53; Gaps 4;
 QY 9 PPYPQRRHAGFSVPLIVLVFLAASFLILPGIRGSRMWLVRYVLSLFTGAEIV 68
 DB 9 PPYAGKPTFPMDTTLASIMFLTALATFVILPGIRGKTRLMVLRVYVLSLFTGAIL 68
 QY 69 AVHFSAEWFGVTNNTSYKAFSAARVTAARVGLVGLGINTLTGTVPVQNLNETIDYNE 128
 DB 69 -----GTPVQNLNETIDYNE 83
 QY 129 QFTWRLEKNYAAYANALEKGLPDPVLYLAKEFTPSSPCGLHYOYHLAGHYASATLMVAF 188
 DB 84 EFTWRLEKNYAAYANALEKGLPDPVLYLAKEFTPSSPCGLHYOYHLAGHYASATLMVAF 143
 QY 189 CFMLSNVLSLTPAPLYGGLALTGTGAVALFGV--FALASISVPLCPRLGSSALFTQY 246
 DB 144 LCMILANVLSMPVLYVGGYMLLATGIFQLALLPFSMATSLTSP-CPHLGASVLTTH 202
 QY 247 GAAPVYTLATGVLCFLGAGVVSIGYRPSALRTLLDQSAK----DCSQERGG 295
 DB 203 GFAFWITLTGTLCLVLLGLAMAVAHRMOPHRLKAFNQSVDDEPMLWSPREGG 256
 RESULT 12
 ABG03520
 ID ABG03520 standard; protein; 445 AA.
 AC ABG03520;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #3511.
 XX
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN W0200175067-A2.
 XX
 PD 11-OCT-2001.
 XX

PF 30-MAR-2001; 2001WO-US008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX
 XX (HSE-) HSE INC.
 PA
 XX
 PI Drmanac RT, Liu C, Tang YT;
 DR WPI: 2001-639362/73.
 DR N-PSDB; AAS67707.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20; SEQ ID NO 33879; 103bp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG03037 represent novel human diagnostic
 CC amino acid sequences. ABG00010-ABG03037 represent novel human diagnostic
 CC amino acid sequences. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 445 AA;
 Query Match 27.6%; Score 460.5; DB 4; Length 445;
 Best Local Similarity 51.1%; Pred. No. 3.6e-41;
 Matches 91; Conservative 14; Mismatches 28; Indels 45; Gaps 1;
 QY 9 PPYPQRRHAGFSVPLIVLVFLAASFLILPGIRGSRMWLVRYVLSLFTGAEIV 68
 DB 77 PPYAGKPTFPMDTTLASIMFLTALATFVILPGIRGKTRLMVLRVYVLSLFTGAIL 136
 QY 69 AVHFSAEWFGVTNNTSYKAFSAARVTAARVGLVGLGINTLTGTVPVQNLNETIDYNE 128
 DB 137 -----GTPVQNLNETIDYNE 151
 QY 129 QFTWRLEKNYAAYANALEKGLPDPVLYLAKEFTPSSPCGLHYOYHLAGHYASATLMV 186
 DB 152 EFTWRLEKNYAAYANALEKGLPDPVLYLAKEFTPSSPCGLHYOYHLAGHYASATLMV 209
 RESULT 13
 AAB66073
 ID AAB66073 standard; protein; 112 AA.
 AC AAB66073;
 XX
 DT 30-MAR-2001 (first entry)
 XX
 DE Human INTERCEPT 296 extracellular domain #2.
 XX
 XX TANGO protein; INTERCEPT protein; neurological disorder;
 KM central nervous system; focal brain disorder; bipolar affective disorder;
 KM global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;

KW senile dementia; Huntington's disease; amyotrophic lateral sclerosis;
 KW Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;
 KW neuropsychiatric; psychoactive substance use; anxiety.
 OS Homo sapiens.
 PN WO200077239-A2.
 XX 21-DEC-2000.
 XX 24-MAY-2000; 2000WO-US014858.
 XX 14-JUN-1999; 99US-00333159.
 XX (MILL-) MILLENNIUM PHARM INC.
 XX McCarthy SA, Fraser CC, Sharp JD, Barnes TM;
 XX WPI; 2001-032313/04.
 XX TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for
 PT screening assays and diagnostic assays and for the treatment of
 PT neurological diseases such as Alzheimer's, Parkinson's and Huntington's
 PT disease.
 XX Claim 8; Page 331; 359pp; English.
 XX The present invention relates to TANGO or INTERCEPT proteins and coding
 CC sequences (see AAF45121-P45136 and AAF45138-P45139 and AAB66031-B66057,
 CC AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding
 CC sequences are useful for the treatment of neurological disorders such as
 CC central nervous system (CNS) disorders, CNS-related disorders, focal
 CC brain disorders, global-diffuse cerebral disorders and other neurological
 CC and cerebrovascular disorders. The CNS disorders include Alzheimer's
 CC disease, senile dementia, Huntington's disease, amyotrophic lateral
 CC sclerosis, Parkinson's, Gilles de la Tourette's syndrome, autonomic
 CC function disorders such as hypertension and sleep disorders,
 CC neuropsychiatric disorders, psychoactive substance use disorders,
 CC anxiety, and bipolar affective disorder.
 XX Sequence 112 AA;
 SQ
 Query Match 27.5%; Score 458; DB 4; Length 112;
 Best Local Similarity 76.8%; Pred. No. 9.3e-42;
 Matches 86; Conservative 10; Mismatches 16; Indels 0; Gaps 0;
 QY 71 HPSAEFVGTNTNTSYKAFSAARVTARVGLVLEGINITLTGPVQALNETIDYNEQF 130
 DB 1 NFSSEWSVGVSTNTSYKAFSSSEWISADIGLVGGVNIITLTGPVQALNETINVEEF 60
 QY 131 TWRKENVAAEYVNALEKGLPDPVLYLAEKFTSPSCGLYHOYHLAGHYASA 182
 DB 61 TWRKENVAAEYVNALEKGLPDPVLYLAEKFTSPSCGLYHOYHLAGHYASA 112
 RESULT 14
 ABO32727
 ID ABO32727 standard; protein; 112 AA.
 XX
 AC ABO32727;
 XX
 DT 17-SEP-2003 (first entry)
 XX
 DE Secreted polypeptide-related protein #101.
 XX
 KW Human; TANGO; INTERCEPT; secreted polypeptide; immune disorder;
 KW hormonal disorder; proliferative disorder; cancer; thyroid disorder;
 KW diabetes; multiple sclerosis; lupus; neurological disorder; anaemia;
 KW Alzheimer's disease; Parkinson's disease; cardiovascular disorder;
 KW myocardial infarction; congestive heart disease; blood platelet disorder;
 KW thrombocytopenia; blood vessel; atherosclerosis; vasculitis.
 XX Homo sapiens.
 OS

XX US2003022279-A1.
 PN 30-JAN-2003.
 XX
 PD 12-JAN-2001; 2001US-00759130.
 XX
 PF 14-JUN-1999; 99US-00333159.
 PR 28-JUN-1999; 99US-00342364.
 PR 10-SEP-1999; 99US-00333996.
 PR 19-OCT-1999; 99US-00420707.
 PR 07-JAN-2000; 2000US-00479249.
 PR 27-APR-2000; 2000US-00559497.
 PR 24-MAY-2000; 2000US-00578063.
 PR 16-JUN-2000; 2000US-00596194.
 PR 23-JUN-2000; 2000US-00602871.
 PR 30-JUN-2000; 2000US-00608452.
 XX
 XX (FRAS/) FRASER C C.
 PA (BARN/) BARNES T M.
 PA (SHAR/) SHARP J D.
 PA (KIRS/) KIRST S J.
 PA (MYER/) MYERS P S.
 PA (LEIB/) LEIBY K R.
 PA (HOLT/) HOLTZMAN D A.
 PA (MCCA/) MCCARTHY S A.
 PA (WRIE/) WRIGHTON N.
 PA (MACK/) MACKAY C R.
 PA (GOOD/) GOODEARL A D J.
 XX
 PI Fraser CC, Barnes TM, Sharp JD, Kirst SJ, Myers PS, Leiby KR;
 PI Holtzman DA, McCarthy SA, Wrighton N, Mackay CK, Goodearl ADJ;
 PI WPI; 2003-456290/43.
 XX
 PT New nucleic acid molecule encoding a secreted protein (e.g. TANGO 202,
 PT TANGO 210 or INTERCEPT 217), useful for diagnosing, preventing or
 PT treating disorders such as cancer, diabetes or atherosclerosis, and in
 PT forensic biology.
 XX
 XX Claim 9; SEQ ID NO 427; 482pp; English.
 PS
 CC The invention relates to secreted polypeptide-related proteins and
 CC nucleic acids (TANGO and INTERCEPT proteins and nucleic acids). The
 CC nucleic acids, proteins and antibodies specific to the proteins are
 CC useful in screening assays, predictive medicine (e.g. diagnostic assays,
 CC prognostic assays, monitoring clinical trials and pharmacogenetics) and
 CC prophylactic and therapeutic methods. The sequences are used in
 CC diagnosing, preventing or treating proliferative disorders (e.g.
 CC cancers), hormonal disorders (e.g. diabetes or thyroid disorders), immune
 CC disorders (e.g. multiple sclerosis or lupus), neurological disorders
 CC (e.g. Alzheimer's disease or Parkinson's disease), cardiovascular
 CC disorders (e.g. myocardial infarction or congestive heart disease), blood
 CC platelet disorders (e.g. thrombocytopenia or anaemia) and disorders
 CC involving blood vessels (e.g. atherosclerosis or vasculitis). The nucleic
 CC acids may also be used in chromosome mapping, tissue typing and forensic
 CC biology, and as surrogate markers. This sequence represents a secreted
 CC polypeptide-related protein of the invention. Note: The sequence data for
 CC this patent was obtained in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html
 XX
 SQ Sequence 112 AA;
 Query Match 27.5%; Score 458; DB 6; Length 112;
 Best Local Similarity 76.8%; Pred. No. 9.3e-42;
 Matches 86; Conservative 10; Mismatches 16; Indels 0; Gaps 0;
 QY 71 HPSAEFVGTNTNTSYKAFSAARVTARVGLVLEGINITLTGPVQALNETIDYNEQF 130
 DB 1 NFSSEWSVGVSTNTSYKAFSSSEWISADIGLVGGVNIITLTGPVQALNETINVEEF 60
 QY 131 TWRKENVAAEYVNALEKGLPDPVLYLAEKFTSPSCGLYHOYHLAGHYASA 182

DB 61 TWRUGENYAECARLEKGLPDPVLYIAEKTPRSPGGLYRQYRLAGHTYSA 112

RESULT 15

AAB66060
ID AAB66060 standard; protein; 383 AA.

XX AAB66060;

DT 30-MAR-2001 (first entry)

DE C06E1.3 related protein, CRP.

XX Gene therapy; TANGO protein; INTERCEPT protein; neurological disorder;
XX central nervous system; focal brain disorder; bipolar affective disorder;
XX global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;
XX senile dementia; Huntington's disease; amyotrophic lateral sclerosis;
XX Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;
XX neuropsychiatric; psychoactive substance use; anxiety.

OS Caenorhabditis elegans.

XX WO200077239-A2.

XX 21-DEC-2000.

XX 24-MAY-2000; 2000WO-US014858.

XX 14-JUN-1999; 99US-00333159.

XX (MILL-) MILLENNIUM PHARM INC.

XX McCarthy SA, Fraser CC, Sharp JD, Barnes TM;

XX MPI; 2001-032313/04.

PT TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for
PT screening assays and diagnostic assays and for the treatment of
PT neurological diseases such as Alzheimer's, Parkinson's and Huntington's
PT disease.

XX Claim 8; Fig 7; 359pp; English.

XX The present invention relates to TANGO or INTERCEPT proteins and coding
XX sequences (see AAF45121-P45136 and AAF45138-P45139 and AAB66032-B66057,
XX AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding
XX sequences are useful for the treatment of neurological disorders such as
XX central nervous system (CNS) disorders, CNS-related disorders, focal
XX brain disorders, global-diffuse cerebral disorders and other neurological
XX and cerebrovascular disorders. The CNS disorders include Alzheimer's
XX disease, senile dementia, Huntington's disease, amyotrophic lateral
XX sclerosis, Parkinson's, Gilles de la Tourette's syndrome, autonomic
XX function disorders such as hypertension and sleep disorders,
XX neuropsychiatric disorders, psychoactive substance use disorders,
XX anxiety, and bipolar affective disorder. The present sequence is a
XX sequence used in a sequence homology alignment with the TANGO/INTERCEPT
XX sequences of the present invention

XX Sequence 383 AA;

Query Match 16.6%; Score 276.5; DB 4; Length 383;

Best Local Similarity 29.9%; Pred. No. 4.2e-21;

Matches 83; Conservative 43; Mismatches 125; Indels 27; Gaps 8;

DB 27 VIVFETALASFLIIPGIRGSRWFMVLVLSLFGAIVAVHPSAEWFGVNTNTS 86

DB 14 IFSVFLPLIAYIILDPGR-RKRVTTTYVYMLAVGALLIASLIPCMASGSMITYQ 72

QY 87 YKAFSAARVAVGLVLEGINITLTGTPHQLNETID-----YNEQFTWLKENY 138

DB 73 FRGHSNERILAKISVEIGLOKVVNTLKEFRLISSNDVLPQSDWTELYYNEGFDISGISM 132

QY 139 AAEYANLLEKGLPPVLYIAEKFT---PSSPCGLHYQYHLAGHYASATLMVAFCMLLSN 195

DB 133 AEAHLHGLENGGLPYPMLSVLEYFSLNQDSFDMG--RHRYVAGHYTHAIIMFAACMLISV 190

QY 196 VL-ISTPAPLYGGLALITTAFAALPGVFALASISSVPLGCLRIGSA-----LTOYG 247

DB 191 VLMLFPHNAV--KSILATGISCLIACLVYLLISP--CELRIFTGENFERVDLTAITS 245

QY 248 AAFVTLATGVLCLFLGAVVSLQYVPPSALRTLLDQS 285

DB 246 FCFYLIFAIGLIVLGLGIGICEHWRITLSTFLDAS 283

Search completed: February 23, 2004, 19:16:58

Job time : 103 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2004, 19:18:06 Search time 44 Seconds
[without alignments]
375.462 Million cell updates/sec

Title: US-09-936-456-2
Perfect score: 1666
Sequence: 1 MTIMGVLPFPQPRHAGF.....GDLHKQALPDKCITTL 320

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Issued Patents, AA*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PTCTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfilest.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	109.5	6.6	457	US-09-543-681A-6481	Sequence 6481, Ap
2	108	6.5	660	US-09-489-039A-8557	Sequence 8557, Ap
3	104	6.2	398	US-09-489-039A-8645	Sequence 8645, Ap
4	99.5	6.0	533	US-09-328-352-7892	Sequence 7892, Ap
5	98.5	5.9	317	US-09-305-856B-12	Sequence 12, Appl
6	98.5	5.9	416	US-09-489-039A-8296	Sequence 8296, Ap
7	97.5	5.9	432	US-08-677-049-8	Sequence 8, Appl
8	97	5.8	413	US-09-489-039A-7562	Sequence 7562, Ap
9	96.5	5.8	486	US-09-252-991A-19571	Sequence 19571, A
10	96	5.8	429	US-09-489-039A-13214	Sequence 13214, A
11	96	5.8	446	US-09-543-681A-6579	Sequence 6579, Ap
12	95	5.7	359	US-08-307-382-2	Sequence 2, Appl
13	95	5.7	359	US-08-366-779-2	Sequence 2, Appl
14	95	5.7	359	US-08-478-727-2	Sequence 2, Appl
15	95	5.7	359	US-08-473-508-2	Sequence 2, Appl
16	95	5.7	359	US-08-789-936-2	Sequence 2, Appl
17	95	5.7	359	US-08-833-610-6	Sequence 6, Appl
18	95	5.7	359	US-08-834-033A-16	Sequence 16, Appl
19	95	5.7	359	US-08-934-254-2	Sequence 2, Appl
20	95	5.7	359	US-09-377-452-6	Sequence 6, Appl
21	95	5.7	359	US-09-685-775-2	Sequence 2, Appl
22	94.5	5.7	403	US-09-328-352-7791	Sequence 7791, Ap
23	94.5	5.7	455	US-09-134-001C-5597	Sequence 5597, Ap
24	94	5.6	1160	US-09-328-352-6826	Sequence 6826, Ap
25	93.5	5.6	569	US-09-252-991A-22870	Sequence 22870, A
26	92.5	5.6	421	US-09-489-039A-7699	Sequence 7699, Ap
27	92.5	5.6	467	US-09-332-041-3	Sequence 3, Appl

28	91.5	5.5	515	US-09-328-352-7312	Sequence 7312, Ap
29	91	5.5	437	US-09-134-001C-4808	Sequence 4808, Ap
30	90	5.4	424	US-09-543-681A-4762	Sequence 4762, Ap
31	90	5.4	503	US-09-252-991A-32663	Sequence 32663, A
32	89.5	5.3	441	US-09-543-681A-6292	Sequence 6292, Ap
33	89	5.3	441	US-09-252-991A-17727	Sequence 17727, A
34	89	5.3	484	US-09-489-039A-7690	Sequence 7690, Ap
35	89	5.3	510	US-09-252-991A-26663	Sequence 26663, A
36	89	5.3	597	US-09-252-991A-24252	Sequence 24252, A
37	89	5.3	405	US-09-252-991A-30503	Sequence 30503, A
38	88.5	5.3	455	US-09-252-991A-23745	Sequence 23745, A
39	87	5.2	475	US-09-328-352-5706	Sequence 5706, Ap
40	87	5.2	482	US-09-489-039A-9909	Sequence 9909, Ap
41	87	5.2	485	US-09-489-039A-9448	Sequence 9448, Ap
42	86.5	5.2	288	US-09-543-681A-5375	Sequence 5375, Ap
43	86.5	5.2	1876	US-09-619-554-2	Sequence 2, Appl
44	86	5.2	450	US-09-252-991A-25019	Sequence 25019, A
45	86	5.2	578	US-09-489-039A-14097	Sequence 14097, A

ALIGNMENTS

RESULT 1
US-09-543-681A-6481: Sequence 6481, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIR
FILE REFERENCE: 2709.1002-001
CURRENT FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US/09/543,681A
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 6481
LENGTH: 457
TYPE: PRT
ORGANISM: Proteus mirabilis
US-09-543-681A-6481

Query Match 6.6% Score 109.5; DB 4; Length 457;
Best Local Similarity 20.2% Pred. No. 0.0055;
Matches 68; Conservative 39; Mismatches 113; Indels 117; Gaps 15;

QY	20	FSVPLIVLVFLALASFLILPGIRGHSRW	FWLVVLLSLFTGAE 66
DB	23	FVSPFHLTFPAITTGILASFLALLEG	MLTKRPSYLSLYQFWIKIFAVNFGVG 77
QY	67	---IYAVFSAEW-----FVGTV-NTNTSYKAPSAAVTA-----	RVGLLVGLE 106
DB	78	SGLVAVYVGFNNWGSFOSFAGSTFGPMLTTEVLTAFLENGFGLGIMLFGNNKVGRLHFF	137
QY	107	GINITLTGTPVHOINETIDYNEQFTWIKEN--YAAEVANALEKGLPDPVLYLAKEFTPS	164
DB	138	ATCWVALTLTLSTF-----WLTSSNMWQTPQYAIENGWVVPVDMQOVVFNPS	186
QY	165	SECGIYHOYHLAGHYASATLMVANC-FWLSNVLSTSPALVYGGMLTTGARALGVFA	223
DB	187	FYYRLHFW-SIGAFNALFTASCAAML-----	214
QY	224	IASISVPLCPRLRSSALTTQGAATVTLATGVLCLPLGAVVSLQYVRPSALRTLLD	283
DB	215	---LKGQATAVR-----KGFMSALMLVILIAPIQAFVGA-----HGLNTLEH	285
QY	284	QSAKDCSOE-----RGSPLITIGDPLHKQALPDL 313	
DB	256	QPAKTAIEGHMDNATKEATPLT-----FGIPDM 285	

RESULT 2

```

Query Match 104; Score 104; DB 4; Length 398;
Best Local Similarity 21.8%; P-Id. No. 0.017;
Matches 71; Conservative 39; Mismatches 99; Indels 122; Gaps 15

QY 12 PQRHNA-----GFSVPLIVLIVLALAAAGLLLPDIRSHSMFLV---RV 57
   | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 135 PEIRGAAYIGRLQALDTVGAFGLFGLVALIMLF--WANDPHAI-----EWANIPAY 183
   | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 58 LLSTFLIG-----AEIVAVHSAEMFVGCTVNTNTSYGASARATYAR 98

```

RESULT 5
US-09-305-856B-12
; Sequence 12, Application US/09305856B
; Patent No. 6479236
; GENERAL INFORMATION:
; APPLICANT: Penny, Laura
; APPLICANT: Galvin, Margaret
; TITLE OF INVENTION: Genotyping the Human
; TITLE OF INVENTION: UDP-Glucuronosyltransferase 1 (UGT1) Gene
; FILE REFERENCE: 4389-7 (formerly SEQ-17C1P)

CURRENT APPLICATION NUMBER: US/09/305,856B
CURRENT FILING DATE: 1999-05-05
PRIOR APPLICATION NUMBER: 60/084,807
PRIOR FILING DATE: 1998-05-07
NUMBER OF SEQ ID NOS: 124
SOFTWARE: PatSeq for Windows Version 3.0
SEQ ID NO 12
LENGTH: 317
TYPE: PRT
ORGANISM: Homo sapiens
US-09-305-856B-12

Query Match 5.9%; Score 98.5; DB 4; Length 317;
Best Local Similarity 22.7%; Pred. No. 0.047;
Matches 62; Conservative 29; Mismatches 77; Indels 103; Gaps 15;

QY 4 WNGVLPFPYQPRHAGSPVLLIVIVFLAASFLILPGIRGSRWFLVRLSLPI 63
DB 6 WTGLPLLY-----VCLLLTCLPSPGKLVVPMDG-SHMTMGSVKXIL 50
QY 64 -GAETVAVHPSAEWFG---TVNT-NTSY-----KAPSAARVTAAGLVGLBG 107
DB 51 RGHVVVVVMEVSWQCRSLNCTKYTSYTLDDREPMVFADARMTAPRSASFSL- 108
QY 108 INTLTGTPVHQL-----NETIDYNEQFTRLKEN-----VAAE 141
DB 109 --LTSSNGIFDLFFSNCRLFNDRKLYEY-----LHSCDVAFLDPFRGLIYAK 159
QY 142 Y-----ANALEKG--LPDFVLYLAEKFTSPSPGGLY-----HOYLAGHYA 180
DB 160 YFSLPSVYFARGFCHYLEGACCPAPLSTYPRLLGLFSDAMFKERVMNHLMHEH-- 217
QY 181 SATLWVAFCEFWLSNVY-----LSTPAPLY 205
DB 218 -----LFCYFFKNVLEIASEILQTVTAY 242

RESULT 6
US-09-489-929A-9296
Sequence 9296, Application US/09489039A
Patent No. 6610835
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
FILE REFERENCE: 2709,2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 9296
LENGTH: 416
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-489-929A-9296

Query Match 5.9%; Score 98.5; DB 4; Length 416;
Best Local Similarity 22.7%; Pred. No. 0.07;
Matches 75; Conservative 50; Mismatches 140; Indels 65; Gaps 16;

QY 17 AAGFVPLILIVFLAASFLILPGIRGSR-WFWLVRVLLSLPI---GAETVAH 71
DB 44 AVGLIGTLFVGLKXAVAVLVLMVMSIANHOGQKTSRPIFLYILGTFSALTAIV 103
QY 72 FSAEMFVGTNTNTSYKASAAVTAAGLVGLGINITLTGTPVHQLNETIDYNEQFT 131
DB 104 FSP-LFSPSTLHLLT-----AASITPPEGIVFVRGLMSVSNIPDALINA-NYIGILV 156
QY 132 WTKENYAAEYNALEKGLPDPVLY-----LAEKTPSPSPGGLYHQ-----YHLAGH 178
DB 157 WAVGIGFALRHGNDTKKLLINDVSHAVTFIVKVIIRFAPLIGIFGLVSTLATTGFEITMG 216

QY 179 YASATLWVAFCEFWLS---NVLL-----STPAPLYGSLALT-----TGAPALFCVAPALA 225
DB 217 YVCLLVIVGCMILVALVINPLVFWKIRPNPPL-----VLTCLRSQGVAFPTRSSAA 271
QY 226 SIS-SVPLC-----PRLGSSALTTGYGAFFWTLATGVLCFLGAVVSLQYVR 274
DB 272 NIPVMALCEKXNLDRDITYSVSIPGATINMAGAAITITVLT-----LAHVHTINIPVDL 326
QY 275 PSALRTLLDQSAKC---SQERGSPPLIL 300
DB 327 PRL--LTSVASTLQCGASGVAGGSLILLI 354

RESULT 7
US-08-677-049-8
Sequence 8, Application US/08677049
Patent No. 5858707
GENERAL INFORMATION:
APPLICANT: Guimaraes, M. Jorge
APPLICANT: Bazan, J. Fernando
APPLICANT: McClanahan, Terrell K.
TITLE OF INVENTION: PURIFIED MAMMALIAN NUCLEOBASE PERMEASES;
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSER: DNAX Research Institute
STREET: 901 California Avenue
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1104

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/677,049
FILING DATE: 03-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/000,788
FILING DATE: 03-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: Chang, Edwin P.
REGISTRATION NUMBER: 34,090
REFERENCE/DOCKET NUMBER: DX0511
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-852-9196
TELEFAX: 415-496-1200
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 432 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Region
LOCATION: 115..144
OTHER INFORMATION: /note= "Encompasses TM 4 of Figure"
FEATURE:
NAME/KEY: Region
LOCATION: 325..359
OTHER INFORMATION: /note= "Encompasses TM 9 of Figure"
FEATURE:
NAME/KEY: Region
LOCATION: 363..390
OTHER INFORMATION: /note= "Encompasses TM 10 of Figure"
OTHER INFORMATION: 4"

[illegible]

NUMBER OF SEQ ID NOS: 14342
PRIOR FILING DATE: 1999-01-29

SEQ ID NO 13214

LENGTH: 429
 TYPE: PRT
 ORGANISM: Klebsiella pneumoniae
 US-09-489-039A-13214

Query Match
 Best Local Similarity 24.2%; Pred. No. 0.13;
 Matches 67; Conservative 35; Mismatches 99; Indels 76; Gaps 14;

QY 21 SVPLIVLVFLALASFLILPGIRGSRWFLVRLV-----LISLFGAIVAHFSAE 75
 DB 59 SRIILAIIVFPAIAPGILMTAPAAQAQASEMOLKQIQGWLFSALGMSLAT-LPAW 117
 QY 76 WFGVTNNTSYKARSARVAVRGVLVGLGEGINITLTGPVHQLNETIYNQFTWRLK 135
 DB 118 WMSRLDWR-----RVALTGIVFLANLVSAYVQY-ETL----- 152
 QY 136 ENYAAEVANALEKGLPDVLYLAKEFTSPSCGLYHQYHLAGHYASATLWVAFCEMILSN 195
 DB 153 --LAARFLASLAGTLMILICISCAAGTP-NPSRYV-----AFWVLQ 191
 QY 156 VILSTPAPLYGSLALTGTGAPALRGVFLASI-SSVPLCPRLGS-----SALTQ 245
 DB 132 LILG---MLGLAL--DGLPATFGLKVVYIILAIMLCLPLVSAPPRFPLVSRRQ 245
 QY 246 YGAAPVTLATGVLCLFLGAVVSLQYVRPSALRTLL 282
 DB 246 PSTALW-RQALAVL-----AVLTF-YISLAVMTFL 274

RESULT 11

US-09-543-681A-6579
 Sequence 6579; Application US/09543681A
 Patent No. 6605709
 GENERAL INFORMATION:
 APPLICANT: GARY BRETON
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
 TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 2709.1002-001
 CURRENT APPLICATION NUMBER: US/09/543,681A
 CURRENT FILING DATE: 2000-04-05
 PRIOR APPLICATION NUMBER: US 60/126,706
 PRIOR FILING DATE: 1999-04-09
 NUMBER OF SEQ ID NOS: 8344
 SEQ ID NO: 6579
 LENGTH: 446
 TYPE: PRT
 ORGANISM: Proteus mirabilis
 US-09-543-681A-6579

Query Match
 Best Local Similarity 20.4%; Pred. No. 0.14; Length 446;
 Matches 61; Conservative 49; Mismatches 117; Indels 72; Gaps 12;

QY 26 IYLVFLALASFLILPGIRGSRWFLVRLVSLFGAEI-VAHFSAEMVGTAVNTN 84
 DB 92 IYALAVISLATVFGVSTPIGL-----LWRFLLGIYIADYPIATSMITEF-----SN 138
 QY 85 TSYKAFSAARVAVRGVLVGLGEGINIT--LTGPVHQLNE----- 122
 DB 139 KKQRAFA-----GFLAAMWYIGATCANLVGLLVDDGWMFLGSAFPCVITLLIG 191
 QY 123 TIDVNECFWRLKENYAAEVANALEKGLPDVLYLAKEFTSPSCGLYHQYHLAGHYASA 182
 DB 192 RDLDPESPLMLIRQRIKCKNMKMLKEGEPVFEADAKTRTELFNKRHSFVLFA 251
 QY 183 TLWVAFCEMILSNVLTSPAPLYGSLALTGTGAPALRGVFLA---SISVPLC----- 233
 DB 252 VIMTC---QVLPMEAIYTFGPQIVGLGMDGRSAAGNVVISLFFMLGCIPIPMFMINOT 308
 QY 234 ---PRLGSSALTQYGAFA-----WTLATGVCLFLGAVVSLQYVRPSAL 278
 DB 309 GRPFLIGSPAMMTL--ALFTLIGIFPNLPIFLVILAPATYAFPGCGCI-LQWLYPNEL 364

RESULT 12

US-08-307-382-2
 Sequence 2; Application US/08307382

Patent No. 5552306
 GENERAL INFORMATION:

APPLICANT: Thomas, Terry L.

APPLICANT: Reddy, Avutu S.

APPLICANT: Nuccio, Michael

APPLICANT: Freysinet, Georges L.

TITLE OF INVENTION: PRODUCTION OF GAMMA LINCENIC

NUMBER OF INVENTION: ACID BY A DELTA 6-DEBUTRASE

CORRESPONDENCE ADDRESS:

ADDRESSER: Scully, Scott, Murphy & Presser

STREET: 400 Garden City Plaza

CITY: Garden City

STATE: New York

COUNTRY: United States

ZIP: 11530

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/307,382

FILING DATE: 14-SEP-1994

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/817,919

FILING DATE: 08-JAN-1992

ATTORNEY/AGENT INFORMATION:

NAME: McNulty, William E.

REGISTRATION NUMBER: 22,606

TELEPHONE: (516) 742-4343

TELEFAX: (516) 742-4366

TELEX: 230 901 SANS UR

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 359 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-307-382-2

Query Match
 Best Local Similarity 21.3%; Pred. No. 0.13; Length 359;
 Matches 61; Conservative 50; Mismatches 101; Indels 74; Gaps 16;

QY 26 IYLVFLALASFLILPGIRGSRWFLVRLVSLFGAEIYVAHFSAEMVGTAVNTN 85
 DB 45 LIIVLMFSAARVAVLPAPV-----FVRLT-----GCVLAIALAA--FSFNGHDA 90
 QY 86 SYKAFSA-ARTVAVRGVLVGLGEGINITLTGTVPVHQLNETIDVNECFWRLKENYAAE-VA 143
 DB 91 NENAVSSPHNRYLVGATYDFVGLS-----SLMRYRNYLHHTT 131
 QY 144 NALBKGLP--DPVLYLAKEFTSPSCGLY--HQYHLAGHYASATLWVAFCEMILSNVLL 198
 DB 132 NILGHDEVIHDDGAVRUSPE--QEHVGIYRFOQFYIMGLY-----LFIPF-YWFLDYDYL 183
 QY 139 STPAPLYGSLALTGTGAF-----ALPGVFLASISSVP-----CRLRGSSALTQ 245
 DB 184 -----VNLKGYHDKIIPFQPLESLALGILKLMGLYFGDLALGFSIPEVL 232
 QY 246 YGAAPVTLATGVLCLFLGAVVSLQYVRPSALRTLLDOSAKDC 289
 DB 233 IGASVYMTYGIYVCTIFMLAHVLESTREFLTPDGESGAIIDDEWAIC 278

US-08-473-508-2

Sequence 2, Application US/08473508
Patent No. 5689050

GENERAL INFORMATION:

APPLICANT: Thomas, Terry L.

APPLICANT: Reddy, Avutu S.

APPLICANT: Nuccio, Michael

APPLICANT: Frevasinet, Georges L.

TITLE OF INVENTION: PRODUCTION OF GAMMA LINOLENIC

ACID BY A DELTA 6-DESATURASE

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSES: Scully, Scott, Murphy & Presser

STREET: 400 Garden City Plaza

CITY: Garden City

STATE: New York

COUNTRY: United States

ZIP: 11530

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/473,508

FILING DATE: 07-JUN-1995

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/307,382

FILING DATE: 14-SEP-1994

APPLICATION NUMBER: US 07/817,919

FILING DATE: 08-JAN-1992

ATTORNEY/AGENT INFORMATION:

NAME: McNulty, William E.

REGISTRATION NUMBER: 22,606

REFERENCE/DOCKET NUMBER: 8383Z

TELECOMMUNICATION INFORMATION:

TELEPHONE: (516) 742-4343

TELEFAX: (516) 742-4366

TELEX: 230 901 SANS UR

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 359 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-473-508-2

Query Match 5.7%; Score 95; DB 1; Length 359;

Best Local Similarity 21.3%; Pred. No. 0.13; Mismatches 101; Indels 74; Gaps 16;

Matches 61; Conservative 50; Mismatches 101; Indels 74; Gaps 16;

```
QY 26 IYIVFLALASFLILPGIRGSHWFWLVVLSLFGAEIVVHFSAEWFGVTNTNT 85
DB 45 LIIVMLFSAMAFVLPVPI-----PPVRL-----GCVVLAIALAA--FSFNVGHDA 90
QY 86 SYKAFSA-ARVTAARVGLVGLGINITLTGTPVHQLNETIDYNEQFTWRLENYAAE-VA 143
DB 91 NNNAYSSNPHINRVGMYDFVGLS-----SFLMKRYRHHYHHTYT 131
QY 144 NALKEKLP---DPVLYLAKEFTPSPCGLY--HQYHLAGHYASATLVNAPCFWILSNVLL 198
DB 133 NTLGHVVEIHGDGAVMSPE--QEHVGIYRFQOQFYTWGLY---LFIPI-YWFLYDVYL 183
QY 199 STPAFLYGLALLTTGAF---ALFGVFLASISSVPL-----CPLRLGSSALTTQ 245
DB 164 -----VLNKGKHDHDKIPFPQLELASLGLTGLMGLGVFGLPLALGFSIPEVL 232
QY 246 YGAAF-WTLATGVLCLF-LGGAIVSLQYVRPSAIRTLIDQSADQC 289
DB 233 IGAIVYMTYGI-VCTIFMLAVLESTEFLLTPDGSAGAIIDEMAIC 278
```

Search completed: February 23, 2004, 19:21:27
Job time : 45 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2004, 19:18:41 ; Search time 626 Seconds
(without alignments)
107,938 Million cell updates/sec

Title: US-09-936-456-2

Perfect score: 1666
Sequence: 1 MTWNGVLPFPQPRHAGF.....GDPHKGALPDLCITTL 320

Scoring table:

BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 809742 seqs, 21153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09C_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10C_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1666	100.0	320	US-10-187-657-1	Sequence 1, Appl1
2	1660	99.6	320	US-10-264-237-2502	Sequence 2502, Ap
3	874	52.5	343	US-09-759-1308-425	Sequence 425, App
4	874	52.5	343	US-10-042-431-55	Sequence 55, Appl
5	697.5	41.9	298	US-10-094-749-2834	Sequence 2834, Ap
6	458	27.5	112	US-09-759-1308-427	Sequence 427, App
7	458	27.5	112	US-10-042-431-57	Sequence 57, Appl
8	276.5	16.6	383	US-09-759-1308-410	Sequence 410, Appl
9	276.5	16.6	383	US-10-042-431-40	Sequence 40, Appl
10	113.5	6.8	792	US-10-156-761-9442	Sequence 9442, Ap
11	104	6.2	652	US-09-815-242-13816	Sequence 13816, A
12	103.5	6.2	502	US-10-156-761-8917	Sequence 8917, Ap
13	103	6.2	98	US-10-106-698-7554	Sequence 7554, Ap
14	103	6.2	460	US-10-369-493-15557	Sequence 15557, A
15	103	6.2	460	US-10-369-493-15930	Sequence 15930, A

16	103	6.2	460	15	US-10-369-493-16298	Sequence 16298, A
17	101	6.1	735	15	US-10-369-493-8829	Sequence 8829, Ap
18	99.5	6.0	477	9	US-09-815-242-13840	Sequence 13840, A
19	98.5	5.9	317	9	US-09-305-8568-12	Sequence 12, Appl
20	98.5	5.9	317	14	US-10-247-159-112	Sequence 12, Appl
21	98.5	5.9	464	14	US-10-238-075-709	Sequence 709, App
22	98	5.9	464	14	US-09-815-242-10144	Sequence 10144, A
23	96	5.8	420	14	US-10-080-170-33	Sequence 33, Appl
24	95	5.7	359	13	US-10-029-786-2	Sequence 2, Appl
25	95	5.7	359	14	US-10-340-779A-12	Sequence 12, Appl
26	95	5.7	359	15	US-10-369-493-2766	Sequence 2766, Ap
27	94.5	5.7	331	9	US-09-738-626-3900	Sequence 3900, Ap
28	94.5	5.7	459	15	US-10-369-493-18271	Sequence 18271, A
29	94	5.6	484	14	US-10-156-761-13878	Sequence 13878, A
30	94	5.6	652	9	US-09-815-242-10449	Sequence 10449, A
31	94	5.6	652	15	US-10-369-493-23658	Sequence 23658, A
32	93.5	5.6	548	14	US-10-289-161A-9	Sequence 9, Appl1
33	93.5	5.6	905	15	US-10-369-493-22577	Sequence 22577, A
34	92.5	5.6	467	9	US-09-887-038-3	Sequence 3, Appl1
35	92.5	5.6	467	10	US-09-828-173-3	Sequence 3, Appl1
36	92.5	5.6	467	14	US-10-310-433-3	Sequence 3, Appl1
37	92.5	5.6	467	15	US-10-310-154-653	Sequence 653, App
38	91.5	5.5	477	11	US-09-833-245-80	Sequence 80, Appl
39	91.5	5.5	477	15	US-10-264-237-2729	Sequence 2729, Ap
40	91.5	5.5	951	14	US-10-276-340-3	Sequence 3, Appl1
41	91.5	5.5	1161	14	US-10-170-102-4	Sequence 4, Appl1
42	91	5.5	306	14	US-10-156-761-12838	Sequence 12838, A
43	91	5.5	371	14	US-10-156-761-14576	Sequence 14576, A
44	91	5.5	469	14	US-10-156-761-12877	Sequence 12877, A
45	90.5	5.4	445	14	US-10-275-554-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1
US-10-187-657-1
Sequence 1, Application US/10187657
Publication No. US20030068311A1
GENERAL INFORMATION:
APPLICANT: Lasek, Amy K.W.
APPLICANT: Baugh, Mariah R.
APPLICANT: Azimzai, Yalga
TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
FILE REFERENCE: PV-0009 CIP
CURRENT APPLICATION NUMBER: US/10/187,657
CURRENT FILING DATE: 2002-07-01
PRIOR APPLICATION NUMBER: PCT/US06/07817
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60/139,565
PRIOR FILING DATE: 1999-06-16
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PERL Program
SEQ ID NO 1
LENGTH: 320
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: incyle ID No. US20030068311A1 4901066CD1
US-10-187-657-1

Query Match 100.0%; Score 1666; DB 14; Length 320;
Best Local Similarity 100.0%; Pred. No. 1.6e-155;
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTWNGVLPFPQPRHAGFVPLLIIVLFALAAFLIILPGIRGSRWFLVAVLLS 60
DB 1 MTWNGVLPFPQPRHAGFVPLLIIVLFALAAFLIILPGIRGSRWFLVAVLLS 60
QY 61 LPIGAETVAHFSSEWVGTVNTSTKASARVAVRVLVGLGINTTLTGPPVOL 120
DB 61 LPIGAETVAHFSSEWVGTVNTSTKASARVAVRVLVGLGINTTLTGPPVOL 120

QY 121 NETIDYNEQFTWMLKENYAAEYANALEKGLPPVLYLAEKFTSPSSPCGLYHQYHLAGHYA 180
Db 121 NETIDYNEQFTWMLKENYAAEYANALEKGLPPVLYLAEKFTSPSSPCGLYHQYHLAGHYA 180
QY 181 SATLWVAFCEWMLSNVLLSTPAPLYGGLALLTTGAFALFGVPAALSSISVPLCPRLGSS 240
Db 181 SATLWVAFCEWMLSNVLLSTPAPLYGGLALLTTGAFALFGVPAALSSISVPLCPRLGSS 240
QY 241 ALTYGYGAFFWVTLATGYLCLFLGGAIVSLQYVRPSALRTLLDOSAKCOSQERGSSPIL 300
Db 241 ALTYGYGAFFWVTLATGYLCLFLGGAIVSLQYVRPSALRTLLDOSAKCOSQERGSSPIL 300
QY 301 GDPPLHKOALPDLKCIITNL 320
Db 301 GDPPLHKOALPDLKCIITNL 320

RESULT 2
US-10-264-237-2502
Sequence 2502, Application US/10264237
Publication No. US20040009491A1
GENERAL INFORMATION:
APPLICANT: Birse et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PA131PI
CURRENT APPLICATION NUMBER: US/10/264,237
CURRENT FILING DATE: 2002-10-04
PRIOR APPLICATION NUMBER: PCT/US01/16450
PRIOR FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: US 60/205,515
PRIOR FILING DATE: 2000-05-19
NUMBER OF SEQ ID NOS: 2876
SOFTWARE: PatentIn Ver. 3.1
SEQ ID NO 2502
LENGTH: 320
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (106)
OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-237-2502

Query Match 99.6%; Score 1660; DB 15; Length 320;
Best Local Similarity 99.7%; Pred. No. 6,3e-155;
Matches 319; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MTLMNGVLPFYPOPRHAGFSVPLLIIVLFLALASFLILPGIRGSRMFWLVRLS 60
Db 1 MTLMNGVLPFYPOPRHAGFSVPLLIIVLFLALASFLILPGIRGSRMFWLVRLS 60
QY 61 LFIGAEIVAHFSAEWVGTNNNTSYKAFSAARVAVRGLVGEINILITGTPVHQL 120
Db 61 LFIGAEIVAHFSAEWVGTNNNTSYKAFSAARVAVRGLVGEINILITGTPVHQL 120
QY 121 NETIDYNEQFTWMLKENYAAEYANALEKGLPPVLYLAEKFTSPSSPCGLYHQYHLAGHYA 180
Db 121 NETIDYNEQFTWMLKENYAAEYANALEKGLPPVLYLAEKFTSPSSPCGLYHQYHLAGHYA 180
QY 181 SATLWVAFCEWMLSNVLLSTPAPLYGGLALLTTGAFALFGVPAALSSISVPLCPRLGSS 240
Db 181 SATLWVAFCEWMLSNVLLSTPAPLYGGLALLTTGAFALFGVPAALSSISVPLCPRLGSS 240
QY 241 ALTYGYGAFFWVTLATGYLCLFLGGAIVSLQYVRPSALRTLLDOSAKCOSQERGSSPIL 300
Db 241 ALTYGYGAFFWVTLATGYLCLFLGGAIVSLQYVRPSALRTLLDOSAKCOSQERGSSPIL 300
QY 301 GDPPLHKOALPDLKCIITNL 320
Db 301 GDPPLHKOALPDLKCIITNL 320

RESULT 3
US-09-759-130B-425
Sequence 425, Application US/09759130B
Publication No. US20030022279A1
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc.
APPLICANT: McCarthy, Sean A
APPLICANT: Fraser, Christopher C
APPLICANT: Sharp, John D
APPLICANT: Barnes, Thomas S
APPLICANT: Kist, Susan J
APPLICANT: Mackay, Charles R
APPLICANT: Leiby, Kevin R
APPLICANT: Wrighton, Nicolas
APPLICANT: Goodheart, Andrew
APPLICANT: Holzman, Douglas A
TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
FILE REFERENCE: MF100-5350MIM
CURRENT APPLICATION NUMBER: US/09/759,130B
CURRENT FILING DATE: 2002-09-15
PRIOR APPLICATION NUMBER: US 09/479,249
PRIOR FILING DATE: 2000-01-07
PRIOR APPLICATION NUMBER: US 09/559,497
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: US 09/578,063
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: US 09/333,159
PRIOR FILING DATE: 1999-06-14
PRIOR APPLICATION NUMBER: US 09/596,194
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US 09/342,364
PRIOR FILING DATE: 1999-06-29
PRIOR APPLICATION NUMBER: US 09/608,452
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/393,996
PRIOR FILING DATE: 1999-09-10
PRIOR APPLICATION NUMBER: US 09/602,871
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: US 09/420,707
PRIOR FILING DATE: 1999-10-19
NUMBER OF SEQ ID NOS: 460
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 425
LENGTH: 343
TYPE: PRT
ORGANISM: Homo sapiens
US-09-759-130B-425
Query Match 52.5%; Score 874; DB 10; Length 343;
Best Local Similarity 58.5%; Pred. No. 1.6e-77;
Matches 112; Conservative 40; Mismatches 74; Indels 8; Gaps 3;
QY 9 PFYPOPRHAGFSVPLLIIVLFLALASFLILPGIRGSRMFWLVRLSFIGAEIV 68
Db 9 PFYPOPRHAGFSVPLLIIVLFLALASFLILPGIRGSRMFWLVRLSFIGAEIV 68
QY 69 AVHPSAEWVGTNNNTSYKAFSAARVAVRGLVGEINILITGTPVHQLNETIDYNE 128
Db 69 AVHPSAEWVGTNNNTSYKAFSAARVAVRGLVGEINILITGTPVHQLNETIDYNE 128
QY 129 QFTWMLKENYAAEYANALEKGLPPVLYLAEKFTSPSSPCGLYHQYHLAGHYASATLWVAF 188
Db 129 QFTWMLKENYAAEYANALEKGLPPVLYLAEKFTSPSSPCGLYHQYHLAGHYASATLWVAF 188
QY 189 CFWMLSNVLLSTPAPLYGGLALLTTGAFALFGV--FALASISVPLCPRLGSSALTYGY 246
Db 189 CFWMLSNVLLSTPAPLYGGLALLTTGAFALFGV--FALASISVPLCPRLGSSALTYGY 246
QY 247 GAFFWVTLATGYLCLFLGGAIVSLQYVRPSALRTLLDOSAK-----DCSQERG 295
Db 247 GAFFWVTLATGYLCLFLGGAIVSLQYVRPSALRTLLDOSAK-----DCSQERG 295

Db 248 GPAFWITLTGGLCVLLGLAMAVAHRMQPHRLKAFNQSVDEDPMLIEMSP EEG 301

```

RESULT 4
US-10-042-431-55
; Sequence 55, Application US/10042431
; Publication No. US20020182675A1
GENERAL INFORMATION:
APPLICANT: MCCARTHY, Sean A
APPLICANT: BARNES, Thomas M
APPLICANT: FRASER, Christopher C
APPLICANT: SHAR, John D
TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING DIAGNOSTIC,
TITLE OF INVENTION: PREVENTIVE, THERAPEUTIC, AND OTHER USES
FILE REFERENCES: 10147-6U2
CURRENT APPLICATION NUMBER: US/10/042,431
CURRENT FILING DATE: 2001-10-25
PRIORITY APPLICATION NUMBER: US 09/333,159
PRIORITY FILING DATE: 1999-06-14
PRIORITY APPLICATION NUMBER: US 09/578,063
PRIORITY FILING DATE: 2000-05-24
NUMBER OF SEQ ID NOS: 79
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 55
LENGTH: 343
TYPE: PRT
ORGANISM: Homo sapiens
US-10-042-431-55

```

Query Match 52.5%; Score 874; DB 13; Length 343;
Best Local Similarity 58.5%; Pred. No. 1.6e-77;
Matches 172; Conservative 40; Mismatches 74; Indels 8; Gaps 3;

```

QY 9 PEPQORHAGVSEVLLIIVLALMASELLJEGISHSMFWLVVLLSTIGAEIV 68
QY 9 PEPQORHAGVSEVLLIIVLALMASELLJEGISHSMFWLVVLLSTIGAEIV 68
Db 9 PFYAGKEPTFPMDDTLTASIMTFLTAFLVILPGIKTKLEFWLNVVTSLTIGAIL 68
QY 69 AVHFSAMFVGIYNINTSYKAFSAKVTARVALLVGBEIGNITTLTGTFVHOINETIDYNE 128
QY 69 AVHFSAMFVGIYNINTSYKAFSAKVTARVALLVGBEIGNITTLTGTFVHOINETIDYNE 128
Db 69 AVHFSAMFVGIYNINTSYKAFSAKVTARVALLVGBEIGNITTLTGTFVHOINETIDYNE 128
QY 129 QENFWLKENYAAAYANALEKGLPDPVLTLAEKFTSSPGGLYHQYHAGHTASATLWAF 188
QY 129 QENFWLKENYAAAYANALEKGLPDPVLTLAEKFTSSPGGLYHQYHAGHTASATLWAF 188
Db 129 QENFWLKENYAAAYANALEKGLPDPVLTLAEKFTSSPGGLYHQYHAGHTASATLWAF 188
QY 129 EFYWRJGENAEBCALTEKGPDPVLTLAEKFTSPSGGLYRQYRLAGHTASATLWAF 188
QY 129 EFYWRJGENAEBCALTEKGPDPVLTLAEKFTSPSGGLYRQYRLAGHTASATLWAF 188
Db 129 EFYWRJGENAEBCALTEKGPDPVLTLAEKFTSPSGGLYRQYRLAGHTASATLWAF 188
QY 189 CFWLLSNVLSTPEAPLYGGLTALTTGAFALFGV--FALASISSVLCPLRYGSALLTQY 246
QY 189 CFWLLSNVLSTPEAPLYGGLTALTTGAFALFGV--FALASISSVLCPLRYGSALLTQY 246
Db 189 LCWLLANVLMSEVLVYGGYMLLATGIFQLTALLPFSMATSITSP-CLHHGASVLTHTH 247
QY 247 GAAFWYTLANGVLCFLGGAUVSLQYVRRSALRTLLDSK-----DQSQRGG 295
QY 247 GAAFWYTLANGVLCFLGGAUVSLQYVRRSALRTLLDSK-----DQSQRGG 295
Db 248 GAAFWYTLANGVLCFLGGAUVSLQYVRRSALRTLLDSK-----DQSQRGG 295

```

RESULT 5
US-10-094-749-2834
/ Sequence 2834, Application US/10094749
/ Publication No. US20030219741A1
/ GENERAL INFORMATION:
/ APPLICANT: ISOGAI, TAKAO
/ APPLICANT: SUGIYAMA, TOMOYASU
/ APPLICANT: OTSUKI, TETSUJI
/ APPLICANT: WAKAMATSU, AI
/ APPLICANT: SATO, HIROYUKI
/ APPLICANT: ISHII, SHIZUKO
/ APPLICANT: YAMAMOTO, JUN-ICHI
/ APPLICANT: ISONO, YUTAKO
/ APPLICANT: HTO, YUKI
/ APPLICANT: OTSUKA, KAORU
/ APPLICANT: NAGAI, KEIICHI
/ APPLICANT: IRIE, RYOTARO
/ APPLICANT: TAMECHIKA, ICHIRO
/ APPLICANT: SEKI, NAOHICO

```

APPLICANT: YOSHIKAWA, TSUTOMU
APPLICANT: OTSUKA, MOTOMYUKI
APPLICANT: NAGAHARA, KENJI
APPLICANT: MASHIRO, YASUHIKO
TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
FILE REFERENCE: 08435/0160
CURRENT APPLICATION NUMBER: US/10/094,749
CURRENT FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 60/350,435
PRIOR FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: JP 2001-328361
PRIOR FILING DATE: 2001-09-14
NUMBER OF SEQ ID NOS: 3381
SOFTWARE: Patentin Ver.. 2.1
SEQ ID NO 2834
LENGTH: 298
TYPE: PRT
ORGANISM: Homo sapiens
US-10-094-749-2834

```

Query Match	41.9%	Score 697.5;	DB 15;	Length 298;
Best Local Similarity	48.6%	Pred. No. 3.1e-60;		
Matches 143;	Conservative 32;	Mismatches 66;	Indels 53;	Gaps 4;

```

QY  PEPQOPRHAAGSVPLLIYIIVLFALAAFLIIPGIRGHRMWLVRVLSIFIGIEIV 68
Db  PFYAGGKPPFPMDDTLASIMIFLTPALATFVIIPGIRGKRLFWLRVVSLEFGAIL 68
QY  AYHFSAMWVGVITNTSTKAFSARVAVRAGLLVLEGITLITGTPVQNLNTIITYNE 128
Db  -----GTPVQNLNTIITYNE 83
QY  129 QPTWRLEKNYAAEYANALAEKGLPDPVLYLAERKTPSSPCGLYHOYHLAAGYAASTLWVAF 188
Db  84 EFTWRIGENYABEYKALKLEKLPDPVLYLAERKTPRSCGLYRQYRLAAGHTSMTLVAF 143
QY  189 CFVLLSNVLSITPAPLYGGLALITTGAAALRGV--FALAASISVPLCPRLRGSALITQY 248
Db  144 LCVLLANVMISWVLYGGIRYMLALAGITQLLALFESNATISLSP--CEYLRGASVYLTHH 203
QY  247 GAAPVYTLATGVLCFLGGAVVSLQYVRPSALRTLLDQSAK----DQSCERGG 295
Db  203 GPAPVYTLITGLCYLLGLANMAVARMQPHRKAFENQSVDEDPMLMSPEEG 256

```

RESULT 6
US-09-759-1308-427
Sequence 427, Application US/097591308
Publication No. US20030022279A1
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc.
APPLICANT: McCarthy, Sean A
APPLICANT: Fraser, Christopher C
APPLICANT: Sharp, John D
APPLICANT: Barnes, Thomas S
APPLICANT: Kirst, Susan J
APPLICANT: Mackay, Charles R
APPLICANT: Myers, Paul S
APPLICANT: Leiby, Kevin R
APPLICANT: Wrighton, Nicolas
APPLICANT: Goodheart, Andrew
APPLICANT: Holtzman, Douglas A
TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
TITLE OF INVENTION: USES.
FILE REFERENCE: MP100-5350NM1M
CURRENT APPLICATION NUMBER: US/09/759,130B
CURRENT FILING DATE: 2002-09-16
PRIOR APPLICATION NUMBER: US 09/479,249
PRIOR FILING DATE: 2000-01-07
PRIOR APPLICATION NUMBER: US 09/559,497
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: US 09/578,063

```

; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/596,194
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/342,364
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/608,452
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/393,996
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: US 09/602,871
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/420,707
; PRIOR FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 427
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-759-130B-427

Query Match
Best Local Similarity 27.5%; Score 458; DB 10; Length 112;
Matches 86; Conservative 10; Mismatches 16; Indels 0; Gaps 0;
```

```

CY 71 HFSAEWVGTVNNTSKAFSAARVAVGLVGLGGINITLTGTPVHQLNETIDYNEOF 130
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 NFSSEWVGTVNNTSKAFSSWISADIGLVGLGVNITLTGTPVQQLNETINNEEF 60
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
131 TWRLKENYAAEVANALEKGLPDPVLYLAKEFTSPSCGLYHOYHLAGHYASA 182
61 TWRLGENYAAECAKALEKGLPDPVLYLAKEFTSPSCGLYROYRLAGHYTSA 112
```

```

RESULT 7
US-10-042-431-57
; Sequence 57, Application US/10042431
; Publication No. US20020182675A1
; GENERAL INFORMATION:
; APPLICANT: MCCARTHY, Sean A
; APPLICANT: BARNES, Thomas M
; APPLICANT: FRASER, Christopher C
; APPLICANT: SHARP, John D
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING DIAGNOSTIC,
; PREVENTIVE, THERAPEUTIC, AND OTHER USES
; FILE REFERENCE: 10147-6U2
; CURRENT APPLICATION NUMBER: US/10/042,431
; PRIOR FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 57
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-042-431-57
```

```

Query Match
Best Local Similarity 27.5%; Score 458; DB 13; Length 112;
Matches 86; Conservative 10; Mismatches 16; Indels 0; Gaps 0;

CY 71 HFSAEWVGTVNNTSKAFSAARVAVGLVGLGGINITLTGTPVHQLNETIDYNEOF 130
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 NFSSEWVGTVNNTSKAFSSWISADIGLVGLGVNITLTGTPVQQLNETINNEEF 60
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CY 131 TWRLKENYAAEVANALEKGLPDPVLYLAKEFTSPSCGLYHOYHLAGHYASA 182
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 TWRLGENYAAECAKALEKGLPDPVLYLAKEFTSPSCGLYROYRLAGHYTSA 112
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```

RESULT 8
US-09-759-130B-410
; Sequence 410, Application US/09759130B
; Publication No. US2003002279A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: McCarthy, Sean A
; APPLICANT: Fraser, Christopher C
; APPLICANT: Sharp, John D
; APPLICANT: Barnes, Thomas S
; APPLICANT: Kiser, Susan J
; APPLICANT: Mackay, Charles R
; APPLICANT: Myers, Paul S
; APPLICANT: Leiby, Kevin R
; APPLICANT: Wrighton, Nicolas
; APPLICANT: Goodheart, Andrew
; APPLICANT: Holtzman, Douglas A
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; PREVENTIVE, THERAPEUTIC, AND OTHER
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
; FILE REFERENCE: MP100-535OMNIM
; CURRENT APPLICATION NUMBER: US/09/759,130B
; PRIOR FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 09/479,249
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/559,497
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/596,194
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/342,364
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/608,452
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/393,996
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: US 09/602,871
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/420,707
; PRIOR FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 410
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-759-130B-410
```

```

Query Match
Best Local Similarity 16.6%; Score 276.5; DB 10; Length 383;
Matches 83; Conservative 43; Mismatches 125; Indels 27; Gaps 8;

CY 27 VILVFLALASFLITPGRGSEWFWLVRLSLFICAEIVAHFSAEWVGTVNNTS 86
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 14 IFSVFLIPLIAYILILIPGR-RKRYVTYTYVTLMAVGALIASIYCMASGOMITYQ 72
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CY 87 YKAFSAARVAVGLVGLGGINITLTGTPVHQLNETID-----YNEQFTWRLKENY 138
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 73 FRHSNERILAKIGVEIGLQKNVTLKRELRILSSNDVPGSDMTLEYNEGFDISGISM 132
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CY 139 AAEYANALEKGLPDPVLYLAKEFT--ESSPCGLYHOYHLAGHYASATLWVAFCEWMLSN 195
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 133 AEAHLHGLNGCLPYPLSLVLEFSLNOSFDWG--RHRYVAGHTHAIIWFAFCACKLSV 190
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CY 196 VL-LSTPAPLYGGLALITGAPALFGVPAIASISSVPLCPRLGSSA-----LTYOG 247
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 191 VLMLFPHNAV--KSLIATGISCLIAVLVILLSP---CEIRIAFTGENFERVDLTATFS 245
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

RESULT 10
US-10-156-761-9442
Sequence 9442, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089

RESULT 11
 US-09-815-242-13816
 Sequence 13816, Application US/09815242
 Patent No. US20020061569A1
 GENERAL INFORMATION:
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Karl L.
 APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 TITLE OF INVENTION: Identification of Essential Genes in
 TITLE OF INVENTION: Prokaryotes
 FILE REFERENCE: ELITRA.011A
 CURRENT APPLICATION NUMBER: US/09/815,242
 CURRENT FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 13916
 LENGTH: 652
 TYPE: PRT
 ORGANISM: Salmonella typhi
 FEATURE:
 NAME/KEY: VARIANT
 LOCATION: (1)...(652)
 OTHER INFORMATION: Xaa = Any Amino Acid
 US-09-815-242-13816

Query Match
 Best Local Similarity 6.2%; Score 104; DB 9; Length 652;
 Matches 63; Conservative 39; Mismatches 97; Indels 88; Gaps 15;

QY 65 AEIVAH-----FSAEFGVTNTSYKAFSARVAVGLVG-----LE 106
 DB 153 ARIGAVHSAVIFGSGSEALAGRIIDSSRLV-----ITADEGVARGRIPAKKVVDALK 207
 QY 107 GINTLTGTPV--HQLNETIDNE--QFTWLKENYAAEYANALEKLP-----D 152
 DB 208 NPNVTSVEHYIVLKRQNDIDQEGRLWVR-----DLIEKASPEHQPEAMNAED 257
 QY 153 PVIYLAKEFTPSSPCGLQYQYLHGH--YASATLWVAF-----CFWLLSNVLLSTAPL 204
 DB 258 PFLIYTGSGTGKPKVLT--TGGVLYAATTFKXVDYHGDLYMCTADV----- 307
 QY 205 YGGLALLTGAFALFGVAFALASISSVLPCLPLGSSALTTQYGAAFVWTLATGVLCFLG 264
 DB 308 ---GWTGHSYLLYG-----FLACGATLMFE--GVNMPPTPAR--MCQVVD 347
 QY 265 GAVSLQYRPSALRTLLDQSAKDCSQERGSPLLI--GDPILHQA 308
 DB 348 KQOVNITYTAPTRALMAEGDKAIEGTDRSSIRLIGSGEPINEPA 394

RESULT 12

US-10-156-761-8917
 Sequence 8917, Application US/10156761
 Publication No. US20030119018A1

GENERAL INFORMATION:
 APPLICANT: OMURA, SATOSHI
 APPLICANT: IKEDA, HARUO
 APPLICANT: ISHIKAWA, JUN
 APPLICANT: HORIKAWA, HIROSHI
 APPLICANT: SHIBA, TADAYOSHI
 APPLICANT: SAKAKI, YOSHIYUKI
 APPLICANT: HATTORI, MASAHIRA
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 FILE REFERENCE: 249-262
 CURRENT APPLICATION NUMBER: US/10/156,761
 CURRENT FILING DATE: 2002-05-29
 PRIOR APPLICATION NUMBER: JP 2001-204089
 PRIOR FILING DATE: 2001-05-30
 PRIOR APPLICATION NUMBER: JP 2001-272697
 PRIOR FILING DATE: 2001-08-02
 NUMBER OF SEQ ID NOS: 15109
 SEQ ID NO 8917
 LENGTH: 502
 TYPE: PRT
 ORGANISM: Streptomyces avermitilis
 US-10-156-761-8917

Query Match
 Best Local Similarity 6.2%; Score 103.5; DB 14; Length 502;
 Matches 73; Conservative 41; Mismatches 102; Indels 137; Gaps 15;

QY 6 GVLFPYQPPRLAAG-----FVPLLIYIVFLALASFLILIGIRGSRMFWLVAV 57
 DB 188 GVLPAVPSRRTKGTGRDWLGAFTLAILV-----LILLISGCH--EKWTSR 235
 QY 58 LLSLFIGAEIVAHFSAEWTVGVNTNTSYK-----AFSAAAVTARVGLVGLGCI 108
 DB 236 TLGAFVGAIVAAV---VWVI---TESRVKEPMVDMRMPVLFMTLAGILLGF--- 284

QY 109 NITLTGPHQINNETIDYNEQFTWLKENYAAEYANALEKLPDPVYLAKEFTPSSPCG 168
 DB 285 -----AMFTQFIQSVLYQ-----MEDV----- 303
 QY 169 LYHQYHLAGYASATLWVAFCEFWLLSNVLLSTAPLYGL-----ALLTGAFALPGV 221
 DB 304 -----AGYFGASVLAASVYLLPPTLVSLVGAQFGVLRIGARVTLAAGACRGVL 356
 QY 222 -----FALAS-----ISSVPL--CPRLGSSALTTQY 247
 DB 357 GFTWLSAHDATASVIGAWTIGLAISFGYASWPNALIVASVPAHQGIANGINSISRSV 416
 QY 248 AAFWTLATGVLCFLGAVVSLQYRPSALRTLLDQSAKDCSQERGSPLLI 300
 DB 417 SA-----VASAVTISLASKTIFL---PDGWPALPDSOPTLSFTIAGAAFVL 461

RESULT 13

US-10-106-698-7554
 Sequence 7554, Application US/10106698
 Publication No. US20030109690A1

GENERAL INFORMATION:
 APPLICANT: Ruben et al.
 TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypep
 FILE REFERENCE: PA00501
 CURRENT APPLICATION NUMBER: US/10/106,698
 PRIOR FILING DATE: 2002-03-27
 PRIOR APPLICATION NUMBER: PCT/US00/26524
 PRIOR FILING DATE: 2000-09-28
 PRIOR APPLICATION NUMBER: US 60/157,137
 PRIOR FILING DATE: 1999-09-29
 PRIOR APPLICATION NUMBER: US 60/163,280
 PRIOR FILING DATE: 1998-11-03
 NUMBER OF SEQ ID NOS: 8564
 SOFTWARE: PatentIn Ver. 3.0
 SEQ ID NO 7554
 LENGTH: 98
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: MISC_FEATURE
 LOCATION: (65)
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 NAME/KEY: MISC_FEATURE
 LOCATION: (85)
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 NAME/KEY: MISC_FEATURE
 LOCATION: (88)
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 NAME/KEY: MISC_FEATURE
 LOCATION: (98)
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 US-10-106-698-7554

Query Match
 Best Local Similarity 6.2%; Score 103; DB 14; Length 98;
 Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 43 PGIGSRMFWLVRVLLSLFIGAEIV 68
 DB 1 PGCLPQRWFWLVRVLLSLFIGAEIV 26

RESULT 14
 US-10-369-493-15557
 Sequence 15557, Application US/10369493
 Publication No. US20030233675A1

GENERAL INFORMATION:
 APPLICANT: Cao, Yongwei
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Goldman, Barry S.

```

; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO: 15557
; LENGTH: 460
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-10-369-493-15557

```

```

Query Match          6.2%; Score 103; DB 15; Length 460;
Best Local Similarity 21.7%; Pred. No. 0.18; Indels 104; Gaps 15;
Matches 69; Conservative 39; Mismatches 106;

```

```

QY 20 FSVPLLIIVFLAIAASFL-----ILPGIRGSR--WF--WLVRLSL----- 61
DB 156 FSLIKVTIIVFIVLGGAAVGMPLADGSPAGVR--HLRADGMFAGHTVPIIMTVAVN 214
QY 62 --FIGAEIVAVHFS-----AEMFYGT-----VNNT 85
DB 215 FAFSGTEILIGIAGETAQAPARAIPLAIRTTILRLVVLFGVTVLVAALLPAHAAVETSP 274
QY 86 SYKAFSAARVATRVGLVLEGINITLTGTPVHQINETIDYNEQFTWRL--KENYAAYA 143
DB 275 FVRAPELLGIPYAAGLL-----NAVILTRILSAANSGLYAAARMLSLNEGTLPARFA 328
QY 144 NALEKGLPDPVLYLAKEFTPSSPCGLYHQYHLAGHYASATLWVA----- 187
DB 329 RLTRRGIPLPALVL-----SMUGLLAL--LTGVYAADTVFAVAISAVSGFAVVVWLSI 380
QY 188 ---FCF---WLSNVLTSTPA-----PLYG--LALITTGAFALFGVPALASISSV 230
DB 381 CASHYCFRRQLRDGIALDSLAYRAPWYPTPLIGALCVLACAGLAFDPQORIALMCGI 440
QY 231 PLCEPLRLGSSALTTOYGA 248
DB 441 PFVALCYGAHAVTORLAA 458

```

```

RESULT 15
US-10-368-493-15930
; Sequence 15930, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO: 15930
; LENGTH: 460
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-10-369-493-15930

```

```

Query Match          6.2%; Score 103; DB 15; Length 460;
Best Local Similarity 21.7%; Pred. No. 0.18;
Matches 69; Conservative 39; Mismatches 106; Indels 104; Gaps 15;
QY 20 FSVPLLIIVFLAIAASFL-----ILPGIRGSR--WF--WLVRLSL----- 61

```

```

DB 156 FSLIKVTIIVFIVLGGAAVGMPLADGSPAGVR--HLRADGMFAGHTVPIIMTVAVN 214
QY 62 --FIGAEIVAVHFS-----AEMFYGT-----VNNT 85
DB 215 FAFSGTEILIGIAGETAQAPARAIPLAIRTTILRLVVLFGVTVLVAALLPAHAAVETSP 274
QY 86 SYKAFSAARVATRVGLVLEGINITLTGTPVHQINETIDYNEQFTWRL--KENYAAYA 143
DB 275 FVRAPELLGIPYAAGLL-----NAVILTRILSAANSGLYAAARMLSLNEGTLPARFA 328
QY 144 NALEKGLPDPVLYLAKEFTPSSPCGLYHQYHLAGHYASATLWVA----- 187
DB 329 RLTRRGIPLPALVL-----SMUGLLAL--LTGVYAADTVFAVAISAVSGFAVVVWLSI 380
QY 188 ---FCF---WLSNVLTSTPA-----PLYG--LALITTGAFALFGVPALASISSV 230
DB 381 CASHYCFRRQLRDGIALDSLAYRAPWYPTPLIGALCVLACAGLAFDPQORIALMCGI 440
QY 231 PLCEPLRLGSSALTTOYGA 248
DB 441 PFVALCYGAHAVTORLAA 458

```

```

Search completed: February 23, 2004, 19:32:06
Job time : 627 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2004, 19:17:08 ; Search time 45 Seconds
(without alignments)
684.029 Million cell updates/sec

Title: US-09-936-456-2

Perfect score: 1666

Sequence: 1 MTLNNGVLFPYQPRHAGF.....GDPHKGALPDKITTLNL 320

Scoring table: BLASTSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	276.5	16.6	383	2	A88533
2	127.5	7.7	391	2	B69808
3	121.5	7.3	401	2	D83760
4	115	6.9	467	2	A81059
5	115	6.9	815	2	T35970
6	113	6.8	509	2	T11043
7	111.5	6.7	381	2	T11546
8	104	6.2	652	2	AB1020
9	104	6.2	703	2	T13393
10	103.5	6.2	456	1	S27616
11	102.5	6.2	322	2	AF2066
12	102	6.1	358	2	G75586
13	101.5	6.1	342	2	D84180
14	101.5	6.1	465	2	AG0648
15	101	6.1	403	2	C83422
16	101	6.1	546	2	AF0571
17	101	6.1	730	2	T44313
18	101	6.1	1074	2	T45094
19	100.5	6.0	381	2	T11312
20	100.5	6.0	414	2	F65097
21	100.5	6.0	414	2	C91125
22	100.5	6.0	414	2	B85970
23	100.5	6.0	414	2	AG0894
24	100	6.0	730	2	S70954
25	99.5	6.0	398	2	G97744
26	99.5	6.0	422	2	AF0072
27	99.5	6.0	536	2	B82559
28	99.5	6.0	958	2	AC0204
29	99	5.9	612	2	G64678

30	99	5.9	734	1	DERZNS	NADH2 dehydrogenas
31	98.5	5.9	378	1	S01190	ubiquinol-cytochro
32	98.5	5.9	379	1	S43267	ubiquinol-cytochro
33	98	5.9	381	1	A53224	ubiquinol-cytochro
34	98	5.9	449	2	S71005	glutamate transpor
35	98	5.9	463	1	GRECNK	nitrite extrusion
36	98	5.9	463	2	H90844	nitrite extrusion
37	98	5.9	463	2	G85702	nitrite extrusion
38	98	5.9	702	2	T13655	NADH2 dehydrogenas
39	97.5	5.9	353	2	T01542	hypothetical prote
40	97.5	5.9	406	2	T43120	conserved hypothet
41	97.5	5.9	432	2	S38893	uracil transport p
42	97.5	5.9	694	2	T13572	NADH2 dehydrogenas
43	97.5	5.9	704	2	T13503	NADH2 dehydrogenas
44	97.5	5.9	706	2	T13391	NADH2 dehydrogenas
45	97	5.8	393	1	CBPOM	ubiquinol-cytochro

ALIGNMENTS

RESULT 1

A88533
Hypothetical protein C06E1.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 02-Nov-2001
C:Accession: A88533
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bi
A:Reference number: A75000; M01D:19069613; PMID:9651916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A:Accession: A88533
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-383 <STC>
A:Cross-references: GB:chr_III; PIDN:AAA27934.1; PID:G289628; GSPDB:GN00021
A:Note: putative
C:Genetics:
A:Map position: 3

Query Match	16.6%	Score 276.5	DB 2	Length 383
Best local similarity	29.8%	Pred. No. 2.2e-15		
Matches	83	Conservative 43	Mismatches 125	Indels 27
				Gaps 8
QY	27	VILVFLALASFLILPGIRGSRMFLVRLSLFIGAELVAVHPSAEWFGVTVNTS	86	
DB	14	IFSVFLPLAVILILPGVR-RKRVTVTVYVLMALVGGALIASLIYPCWASGQWIVTQ	72	
QY	87	YKASARVAVRPGVLVGLGINTLTGTPIVQNTID-----VNEQFTWRLKENY	138	
DB	73	FRGHSNRILAKIGVEIGQKAVNTLKFRLSSNVVLGSPMTBYVNEGDISGISM	132	
QY	139	AAEYANALEKGLDPVLYLAKEFT---PSPQGLYQYHAGHYASATLVAFCEMLLSN	195	
DB	133	AEALHQLENGILPVMLSVLEYFSLNQDSFDWG--RHYVAGHYTAALWFAFCMCJSV	190	
QY	196	VL-ISTPAPLYGGLALITTAFAALRGVAFALASISSVPLPLRLGSSA-----LITQVG	247	
DB	191	VLMFLPHNAV--KSLIATGISCLIAVLVLLISP---CELRIFAIGENFERVDLATPS	245	
QY	248	AAFWVLATGVLCFLGAVVSSQVYRPSALRTLLDQS	285	
DB	246	FCFYLIFALIGLCVLCGLGICEHRRITLSTFLDAS	283	

RESULT 2

B69808
Multidrug-efflux transporter homolog ykfF - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C:Accession: B69808

R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertet, C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capano, V.; Carter, N.M.; Chd A.; Ehrlich, S.D.; Emerson, P.T.; Ehtian, K.D.; Errington, J.; Fabeet, C.; Ferreri, E. Nature 390, 299-356, 1997

A;Authors: Follger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.F. Koether, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauei, Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelli, Riger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schrotter, R.; Scofield, F.; Sekiguchi, J.; Sekowska, A.; Serot, Akenchi, M.; Tamakoshi, A.; Tanaka, T.; Terpestra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K. A;Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A;Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. A;Reference number: A69580; MUID:98044033; PMID:9884377

A;Accession: B69808

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-391 <KUN>

A;Cross-references: GB:299108; GB:AL009126; NID:q2633055; PIDN:CAH12620.1; PID:e1182781; A;Experimental source: strain 168

C;Genetics:

A;Gene: yfkF

Query Match 7.7%; Score 127.5; DB 2; Length 391;
Best Local Similarity 25.3%; Pred. No. 0.0053;
Matches 75; Conservative 33; Mismatches 105; Indels 83; Gaps 16;

Db 3 LKNGVL---PFYQPRHAGSVPLLIIV---ILVFLALASPLLIIPGIRGHRMFWLRV 57
48 LKGVLASPFMEPLKLGFK-PLIVMGSIYILSRFPIVL-----QSVWVF--- 96
Oy 58 LLSIFIAEIVAAHFAEMFVGVNTNTSYKASAAVTAAGVGLVGLGINITLTGTPV 117
Db 97 LRLFLIGIDHMLHFSQTQVW---TSMSSKQ-NGRNLSIYGLSGFL-GFAAGFPMPVL 150
Oy 118 HONNETIDY-----NEQFTWRLKENYAAEYANALEKGLDPLVLAKEFTPS- 166
Db 151 VKLSPLPFIYVSGISIFANLF-----VFLLONAIPEFTSPHETSD 191
Oy 167 --CGLYHQYHLAGHVASATLWVAFCEFWLLSNVLLSTPAPLVGGLALTTGAFALFGVPAL 224
Db 192 NSFRFYQAMLFQ-----WVAF-----NPTGYGLLETALNGSPVYALRLG 233
Oy 225 ASISVSP--LCPRLGS-----SALTTOYGA-APWTLATGVICLFGAVVS 269
Db 234 ISVDVAAILLPAPALGSIIFQFPLGILSDKXGRNVLLVLTGALCFPIAGVPS 289

RESULT 3
DB3760
hypothetical protein BH0884 [imported] - *Bacillus halodurans* (strain C-125)
C;Species: *Bacillus halodurans*
C;Date: 01-Dec-2000 #sequence revision 01-Dec-2000 #ext_change 15-Jun-2001
C;Accession: DB3760
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: DB3760
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-401 <STO>
A;Cross-references: GB:AP001510; GB:BA000004; NID:g10173440; PIDN:BAB04603.1; GSPDB:GN00 C;Genetics:
A;Experimental source: strain C-125
A;Gene: BH0884

Query Match 7.3%; Score 121.5; DB 2; Length 401;
Best Local Similarity 24.4%; Pred. No. 0.017;
Matches 73; Conservative 45; Mismatches 94; Indels 87; Gaps 19;

Oy 3 LKNGVL---PFYQPRHAGSVPLLIIVFLALASPLLIIPGIRGHRMFWLRV 59
Db 56 LKGVLASPFMEPLKLGFK-PLIVMGSIYILSRFPIVL-----QSVWVF--- 96
Oy 60 SLFPAEIVAAHFAEMFVGVNTNTSYKASAAVTAAGVGLVGLGINITLTGTPV 113
Db 107 RMIGIAQNMVFAFQVWITTSMPKKGRNISFYGA-----FGIGGLGPPM--- 155
Oy 114 GRYVHONNETIDY-----NEQFTW---RLKENYAAEYANALEKGLDPLVLAKEFTPS 164
Db 156 -TRLQINEFLPFIISATSPAMLLIRLNEYPAQ---DVEYG----- 196
Oy 165 SPQGLYHQYHLAGHVASATLWVAFCEFWLLSNVLLSTPAPLVG---GLAL---LTTGAPA 217
Db 197 SOSGWTWRYKAVVKKGMFALLPAFCYGLJESSLGN-FPYVGLSGGLTVEQVSIILPAFV 255
Oy 218 LFGVFAIASISSVPLCP--FLGSSALTYGCAFMTWLTANGVLCFLGAV-VSLOVY 273
Db 256 VGGV---TQMPFLGFLSDKGRKPL-----LWVLFEGATFVSMVYL 295

RESULT 4
AG1059
probable membrane protein STY4802 [imported] - *Salmonella enterica* subsp. *enterica* s. *enterica*
C;Species: *Salmonella enterica* subsp. *enterica* serovar *Typhi*
A;Note: this species has also been called *Salmonella typhi*
C;Date: 09-Nov-2001 #sequence revision 09-Nov-2001 #ext_change 18-Nov-2002
C;Accession: AG1059
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Fair S.; Moutie, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, A.; Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* s. *enterica* serovar *Typhi* strain 04/04/01
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AG1059
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-467 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD06924.1; PTD:g16505572; GSPDB:GN00176 C;Genetics:
A;Gene: STY4802
C;Superfamily: Haemophilus influenzae conserved hypothetical protein HI0594

Query Match 6.9%; Score 115; DB 2; Length 467;
Best Local Similarity 24.2%; Pred. No. 0.07; Mismatches 75; Indels 94; Gaps 11;
Matches 63; Conservative 28;

Db 27 VILVFLALASPLLIIPGIRGHRMFWLRVLLSLFIAEIVAAHFAEMFVGVNTNTS 86
Db 264 VLIIF---AASFVWLYGVAVLGWMN-----AEISGVFLAAIIVGYI-TRMG 307
Oy 87 YKAFSA-----ARVTRVGLVGL-EGT-----NITLTGTVHONNETIDYNEQFTWRLK 135
Db 308 EEAFSTFTIDGADLLIGVALLTIGIARGVYVMDNMITHLLHS----- 351
Oy 136 ENYAAYANALEKGLDPLVLAKEFTSPQGLYHQYHLAGHVASATLWVAFCEFWLLSN 195
Db 352 -----AASIVSGL-----STTIFINVTYML-E 372
Oy 196 VILSTAPLVGGLALTTGAPLFGVPAIASISSVPLCPRLGSSALTYGCAFMTWLTANGVLCFLGAV 255
Db 373 VILSPVFPSSGGLAVITPIMPPLADFAHVORDIV-----VTAQASAGIVNLV 421
Oy 256 TGVLCFLG-AYVSLQVYR 274
Db 422 TPTSAYVWGLAIPVYR 441

RESULT 5
T35970
probable efflux protein - *Streptomyces coelicolor*
C;Species: *Streptomyces coelicolor*

C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C/Accession: T15970
 R/Seeger, K.J.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, January 1999
 A/Reference number: Z21551
 A/Accession: T15970
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-815 <SEE>
 A/Cross-references: EMBL:AL035161; PIDN:CAA22731.1; GSPDB:GN00070; SCODEB:SC9C7.19
 A/Experimental source: strain A3(2)
 C/Genetics:
 A/Gene: SCODEB:SC9C7.19

Query Match 6.9%; Score 115; DB 2; Length 815;
 Best Local Similarity 24.2%; Pred. No. 0.13;
 Matches 85; Conservative 25; Mismatches 107; Indels 134; Gaps 15;

Cy 12 PQPHAGFSVPLIVLPLALAAFLILPGIRGSRFWLVRLSLFTGAE----- 66
 Db 226 PRPR-----VKPRDLIGALLAAATCLVLTWSGTEYAMGSRVILGIAAQAATLL 280
 Cy 67 -IVAVHFSAEWFGVTNTNTSYKAF--SAARVAVGLVGLGIVITLTGTPVHQLNET 123
 Db 281 FLVAHFAPPELI-----PLRFRDSINVTALVGLVVG-----VALFG----- 319
 Cy 124 IDYNEQFTRLKENYAAEY-----ANALEKGLPDPVLYLAEKTPS----- 164
 Db 320 -----AASVLPTELOWVDGASATESG-----LMLPMGQVVGAS 355
 Cy 165 -----SPCGLYHQYHLAGHVASATLVAFCEFWLLSNVLSLTPAPLYG--GLALLTTGANA 217
 Db 356 ISGQISRTGTHYRHPILGSAISV-----GMWLSRLDADTSRLHYSIWVAVLGAGIGL 410
 Cy 218 LFGVFALASISVPLCEPLRGSSALTTQY-----GAAFWTLATGVLCPLGGA VVS 269
 Db 411 VMPPLVAVQNSVR--PTDGTATSNANNYFRQIGSGAAVFGTLFAG----- 456
 Cy 270 LQYVPSALRTLQDSKDCSGERGSPLLDGPHKQALPRLKITTNL 320
 Db 457 -----RLTALADRIPEAG-----VGLPDAEAITPOL 484

RESULT 6
 T11043
 Cytochrome-c oxidase (EC 1.9.3.1) chain I - Chlamydomonas eugametos mitochondrion
 C/Species: mitochondrion Chlamydomonas eugametos
 C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 11-May-2000
 C/Accession: T11043
 R/Denovan-Wright, E.M.; Nedelcu, A.M.; Lee, R.W.
 Plant Mol. Biol. 36, 285-295, 1998
 A/Title: Complete sequence of the mitochondrial DNA of Chlamydomonas eugametos.
 A/Reference number: Z17244; MUID:98145434; PMID:9484440
 A/Accession: T11043
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-509 <DEN>
 A/Cross-references: EMBL:AF008237; NID:G2865253; PID:G2865259; PIDN:AAC39342.1
 C/Genetics:
 A/Genome: mitochondrion
 A/Introns: 365/3
 A/Note: cox1
 C/Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
 C/Keywords: copper; electron transfer; heme; iron; magnesium; membrane-associated complex
 F/10-254/Domain: cytochrome-c oxidase chain I homology <COI>
 F/62,374/Binding site: heme a iron (His) (axial ligands) #status predicted
 F/237,226,287/Binding site: copper (His) #status predicted
 F/237,224/Cross-link: 1-histidyl-3-cysteine (His-Tyr) #status predicted
 F/241/Binding site: oxygen (Tyr) #status predicted
 F/364/Binding site: magnesium (His) (shared with chain II) #status predicted
 F/372/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 6.8%; Score 113; DB 2; Length 509;

Best Local Similarity 21.2%; Pred. No. 0.11;
 Matches 66; Conservative 38; Mismatches 118; Indels 90; Gaps 14;

Cy 26 IVILVFALAAASFLILPGIRGSRW-----FWLVRLSLFTGAE 65
 Db 64 IIMLFVMPALF-----GGGFWLVPLILGADPMAPRLNLSFWLNPALGLL 116
 Cy 66 EIVAVHFSAEWFGVTNTNTSYKAFSAARVAVGLV-----GLE-----GINITLTGT 115
 Db 117 TWVQAGAGTCW-----TAVPUSIGTGAAVDALISLHNLSSILSINILVITA 168
 Cy 116 PVHQLNETIDYNEQFTRLKENYAAEYANALEKGLPDPVLYA-----EKTTP 163
 Db 169 GMRPAGIKLSQMPDFWMSIA-----FTAILVILAVPYLAALVMTDRNLNTAYCE 221
 Cy 164 SSPCGLY-HQYHLAGHVASATLVAFCEFWLLSNVLSLTPAPLYGALL--TTGAFALFG 220
 Db 222 SGDLILYQHLFWFGH--PEVYILILPAGIVSHVISFPGKPIFGNMKICMGALISILG 280
 Cy 221 VFALAS-----ISSVPLCEPLRGSS--SALTTOYGAARFVT---LATGV 258
 Db 281 FIVAAHMFVVGJDLDTIAYFVSATWIIAVPTGMKIFSWLATITYGSLMLTTPMFAVGF 340
 Cy 259 LCLPLGGA VVS 270
 Db 341 ICLFTLGGVTVG 352

RESULT 7
 T11546
 Ubiquinol-cytochrome-c reductase (EC 1.10.2.2) cytochrome b - spiny dogfish mitochondrion
 C/Species: mitochondrion Squalus acanthias (spiny dogfish)
 C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 03-Jun-2002
 C/Accession: T11546
 R/Rasmussen, A.S.; Arnason, U.
 J. Mol. Evol. 48, 118-123, 1999
 A/Title: Phylogenetic studies of complete mitochondrial DNA molecules place cartilag.
 A/Reference number: Z17281; MUID:99091711; PMID:9873084
 A/Accession: T11546
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-381 <RAS>
 A/Cross-references: EMBL:Y18134; NID:G4186095; PIDN:CAA77061.1; PID:G4186108
 C/Genetics:
 A/Genome: mitochondrion
 C/Superfamily: cytochrome b; cytochrome b homology; cytochrome b6 homology; plastocyanin
 C/Keywords: electron transfer; heme; iron; metalloprotein; mitochondrion; oxidative I
 F/112-340/Domain: cytochrome b6 homology <CYB>
 F/112-210/Domain: cytochrome b6 homology <CYB>
 F/122-340/Domain: plastocyanin-plastocyanin reductase 17K protein homology <17K>
 F/184,183/Binding site: heme iron (His) (axial ligands) (low potential) #status predicted
 F/98,197/Binding site: heme iron (His) (axial ligands) (high potential) #status predicted

Query Match 6.7%; Score 111.5; DB 2; Length 381;
 Best Local Similarity 25.8%; Pred. No. 0.11;
 Matches 70; Conservative 35; Mismatches 103; Indels 63; Gaps 14;

Cy 27 VIILVFALAAASFLILPGIRGSRW-----FWLVRLSLFTGAE 76
 Db 118 VILFLMATFAFGVLP-----WGQSFAGATVITNLASAPYIGMLT-----QW 164
 Cy 77 FVGTVN-----TNTSYKAFSAARVAVGLV-LVGLGIVITLTGTPVHQLNETID----- 125
 Db 165 IMWGFSDNATLRFAPFHLPLVGLTTLHLFLHRTGSNNM--GINSMDKISFHP 223
 Cy 126 ---YNEQFTRLKENYAAEYANALEKGLPDPVLYLAEKTPSSPGGLVHQYHLAGHVASA 182
 Db 224 YFSKDLGFLMIIILALALFLPNTLDB-----AENFTPNP--LVTPPIKEW--- 273
 Cy 183 TLWAFCEFWLLSNVLSLTPAPLYGGLALTGAFALFGFALASISVPLCEPLRGSSAL 242
 Db 274 -----YFLFAVAIAMSINKKCGVAL-----LFSIFILMLPMHTSKQSRNIRP 320

QY 243 TTGYGAFFWTATATGCLFPGGAVVSLQV 273
 DB 321 MTQF--LFWTLVANNALITLWIGQVPEQPF 349

RESULT 8

AB1020

acetic-CoA ligase (EC 6.2.1.1) - *Salmonella enterica* subsp. *enterica* serovar Typhi (str C/Species: *Salmonella enterica* subsp. *enterica* serovar Typhi

A/Note: this species has also been called *Salmonella typhi*

C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C/Accession: AB1020

R/Parikh, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, T.; Conerton, P.; Croftin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moutie, S.; O'Garra, P.

Nature 413, 848-852, 2001

A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.

A/Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov

A/Reference number: AB0502; NCID:21534947; PMID:11677608

A/Accession: AB1020

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1652 <PAR>

A/Cross-references: GB:AL513382; PIDN:CAD09259.1; PID:g16505263; GSPDB:GN00176

A/Genetics:

C/Superfamily: acetate-CoA ligase; acetate-CoA ligase homology

C/Keywords: acid-thiol ligase; coenzyme A

Query Match 6.2%; Score 104; DB 2; Length 652;

Best Local Similarity 22.0%; Pred. No. 0.84;

Matches 63; Conservative 39; Mismatches 97; Indels 88; Gaps 15;

QY 65 AETVAVH-----FSAEMFVGVNTNTSYKAFSAARVPAVRLVGLG-----LE 106

DB 153 ARIGAVHVIFFGFSFEALAGRIDDSSRLV-----ITDDEVVRGRITPLKXVDDMLK 207

QY 107 GINTLTGTPV--HQNETIDYNE--QFTWRLKENVAEYANALEKGLP-----D 152

DB 208 NPNVTSEHYIVLAKRGNDIDWQEGRDLMWR-----DLIEKASPEHQPEAMNED 257

QY 153 PVLVAEKFTPSSPCGLYHQYLAGH--YASATLWVAF-----CFWLSNVLLSTPAVL 204

DB 258 PLFILTSGSTGKPKVLT--TGGTLVYATTFKVPDHYHGDLYMCTADV----- 307

QY 205 YGGLALLTTGAPALFVFPALASISSVPLCPRLGSSALTTQYGAFFWTATATGCLFPG 264

DB 308 ---GWTGHSYLLYG-----PLACGATTLMEF--GVNMPPTPAR--MCQVVD 347

QY 265 GAVVSLQVVRPSALRTLLDOSAKDCSQERGSPLL---GDPLHQA 308

DB 348 KQGVNITLYTAFTAIRALMAEGDKAIGTTRSSURLIGSVGEFINDEA 394

RESULT 9

T13393

NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - *Lichachne pauciflora* chloroplast

C/Species: chloroplast *Lichachne pauciflora*

C/Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 03-Jun-2002

C/Accession: T13393

R/Clark, L.G.; Zhang, W.; Wendel, J.F.

Syst. Bot. 20, 436-460, 1995

A/Title: A phylogeny of the grass family (Poaceae) based on ndhF-sequence data.

A/Reference number: Z17570

A/Accession: T13393

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-703 <CLA>

A/Cross-references: EMBL:U21978; NID:g755846; PID:g755847; PIDN:AAA64697.1

A/Experimental source: leaf

C/Genetics:

A/Genome: chloroplast

A/Note: ndhF
 C/Superfamily: NADH dehydrogenase (ubiquinone) chain 5
 C/Keywords: chloroplast; membrane-associated complex; NAD; oxidoreductase

Query Match

Best Local Similarity 6.2%; Score 104; DB 2; Length 703;

Matches 68; Conservative 47; Mismatches 102; Indels 86; Gaps 18;

QY 23 PLITVILFPLAASFLI-----LPGRGSRFMYLVY---LISLFGAEIVAHFS 73

DB 82 PLTSMILLITVIGLILVSDNYSMDHGLRFFVVISFFNTSMGLVSSMLQIYFF 141

QY 74 AE-----WFGVTNNTSYKAFSAARVTAAY---GLLVGEGINTLTGTPVH 118

DB 142 WELVWMCYVLLIGFFFTALIAASACQARF---VTNNVSGDFGLLGLIGF--FWITG--- 132

QY 119 QNETIDYNEQFTWRLKENYAE-----YANALEKGLPDV--LYLAKEFT 162

DB 193 ---SLEPRDLF--KIANNW-PDNGINSLTTLCAFLIFGAVAKSAQFPLHMLPDAME 246

QY 163 PSP-CGYHQYHLAGHYASATLWVAFCFWLSNVLLSTPAPIYGGALTTGAFALF-- 219

DB 247 GFTPLSALIH-----AATWAAAGFLLARFPLFVPLPLMSVSL--VGTITLFLG 236

QY 220 GVPAIA-----SISVPLCPRLGSSALTTQYGAFF--VTLATGVLCLPIG-G 265

DB 297 ATLALAQDRIKRSILAVSTMSQIGYMWLALGIGS---YQALFPLITHAVSKALLFLGSG 352

QY 266 AVV 268

DB 353 SII 355

RESULT 10

S27616

probable glucarate transporter - *Pseudomonas putida*

C/Species: *Pseudomonas putida*

C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C/Accession: S27616

R/Burlingame, R.P.; Maruya, A.; Ally, A.; Ally, D.; Backman, K.C.

submitted to the EMBL Data Library, June 1992

A/Description: Nucleotide sequences of hydroxyproline-specific alpha-ketoglutarate s

A/Reference number: S27612

A/Accession: S27616

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-456 <BUR>

A/Cross-references: EMBL:M69160; NID:g151313; PIDN:AAA25867.1; PID:g151314

C/Superfamily: hexuronate transporter

C/Keywords: membrane protein

Query Match 6.2%; Score 103.5; DB 1; Length 456;

Best Local Similarity 22.8%; Pred. No. 0.62; Indels 125; Gaps 18;

Matches 80; Conservative 35; Mismatches 111; Indels 125; Gaps 18;

QY 19 GFSVPLIIVL---VFALASFLILPGIRGSRFWLVRLSLFGAEIVAHFSA 74

DB 101 GSTAVVLLFLRFPVGAABPSF---PG-----NARIV-----A 132

QY 75 EMFVGVNTNTSYKAFSAAR--VTAHVGLVGLGINTLT-----GTVHQNETIDY 126

DB 133 SMF--PTKRGTSALFNSAQYFAITRAVALDGLDRHLRLAARVRRHGRGHCVLAHVD 191

QY 127 NEQFTWRLKENYAEYANALEKGLPD-----PVL--YLAKEFTPSSPCGLYHQY 174

DB 192 GDLRAERSPAGYAAVRSRPHGLVLEDSKDKDGPKMYITQLLTNNMNGIY---- 247

QY 175 LAGHYASATLWVAFCFWLSNVLLSTPAPIY---GGALITTGAPF-----L 218

DB 248 -LGQFCINALTIVFFLTWF-----PVLVQERGGTITIKAGIISLPAICGFLGVL 236

QY 219 FGVF-----ALASISSVPL--CPRLGSSALTTQYGAFFWTATATGCLF----- 262

D6 297 GGTISDTLLRSGSLVAKRTPIVCGNVLMSENIICNVYDDPMVVCMLAFFGKALGA 356

D5 263 LGGAAVSLQVVRPSALRTLLDQSAKDCQSGERG-----SPILIG 301

D5 357 LGGAAVVS-----DTSPQKIDLSGGLFNFTFGNLSISTPILIG 394

RESULT 11

hypothetical protein all12084 [imported] - Nostoc sp. (strain PCC 7120)
C/Species: Nostoc sp. PCC 7120
A/Name: Nostoc sp. #strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C/Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
R/Accession: AF2066
R/Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchin,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, W.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anaba
A/Reference number: AB1807; MUID:21595285; PMID:11759840
A/Accession: AF2066
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-322 <KUR>
A/Cross-references: GB:BA000019; PTDN:BA873783.1; PID:g17131175; GSPDB:GN00179
A/Experimental source: strain PCC 7120
C/Genetics:
/Gene: all12084

Query Match	6.2%;	Score 102.5;	DB 2;	Length 322;
Best Local Similarity	24.1%;	Pred. No. 0.5;		
Matches 77;	Conservative 43;	Mismatches 116;	Indels 83;	Gaps 16

QY 3 LMGVLP-----FYQPPHAAAGSFYPLIVL-----VFLLAASFLIIILPGIGSHRWF 52
 Dd 6 LMAVLPPIVFLGYYYRVAPHAPPLSLTLMFLGLAIGSFLLGLEIYV-----DSVAN 60
 QY 53 WLNV---RVLSLSF-----IGAELVAHVSAMFVGTON-TNTSYAFSAAR 94
 Dd 61 WYLMWQRIQRLSFLGIALRQIIAIGPIEBECKLAVAVANYFYFORRYHITSSISLFTIAA 120
 QY 95 V---TAR-----VGLLVLEGINITLTGTVPQHONETIDNBQFTRLKENYAA-----EYA 143
 Dd 121 ALGFTAEEMNIYIYHSTATILIDRSIGTPHAA-----FSAPWGVALGKYFPAITRLNHYR 175
 QY 144 NAEKQJPDDEVLY-----LAEKFTPSBPGLYHOYHLAHGYSATLWVAFCW---LL 193
 Dd 176 NLLVRGINSIVFHLVNLVSAMRSPLEP-----LSYGFPLLVWM---FWRMEQL 227
 QY 194 SNVILSTPAPLVGG-----TALTGTGAFALFGVAFALASISVPLCLPRGSS 240
 Dd 228 RQYTSQRPLNLSIGHNLSLRWQRGVLFALMLGQNALFGLEFLARVIS-PLSPAQLFSP 286
 QY 241 ALTYQGAFAWYTLATCVL 259
 Dd 287 DFFWLIMSRFLTNLAEGIL 305

RESULT 12

uree/short chain-mide ABC transporter, permease protein - *Deinococcus radiodurans* (strain ATCC 49239) [10]
 C/Species: *Deinococcus radiodurans*
 C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
 C/Accession: G75586
 R/White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
 S.M.; Shen, M.; Vamshaveyan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A/Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.
 A/Reference number: A75250; MUID:20036866; PMID:10567266
 A/Accession: G75586
 A/Status: Preliminary
 A/Molecule type: DNA
 A/Residues: 1-358 <WHI>

A;Cross-references: GB:AE001863; GB:AE001825; NID:96460670; PIDN:AAF12457.1; PID:964
A;Experimental source: strain R1
C;Genetics:
A;Gene: DRA032
A;Map position: 2
C;Superfamily: 1-arabinose transport system permease arah

Query Ma
Best 1.00

	Matches	66, Conservative	26, Mismatches	98, Indels	44, Gaps	11,
Cy	3	LMNGV--LPFYQCPHAGSFSVPLIIVLVLAALAAAFLLITGICGHSNFWLVKYLIS	60			
Dh	94	LYNGVEKLPFWFQCPHSAFPALLVWVA.PGLIAALITGLIMFRRRTG-----VFVS	144			
Cy	61	LPFGAIVVAHVSAAEMFVG---TVNTV--TSYK-----AFSAATVTARVCL-V	103			
Dh	145	IITQALVLAFA---VWLSGAAGLTSGNNGIIDDPTFLGLMMRSDAFARGLYWMTVLLAL	201			
Cy	104	GLEGINITLTGTPVHQLNETIDYNEQFTWIKENYAAEVANALEKGI.PDVLVLAKEFTP	163			
Dh	202	SLAG-TAWMLLRPFSGSLITAAVDNENNRFLGYNPAPAFKVAALAVLG---GVL-----	249			
Cy	164	SSPCGLHYOHYLAHYASATLVVAFCFWLLSNVLTSTPAFLYGGALLTTTGARA	217			
Dh	250	SGISGALYLTIH-GTISPMITGVAFSFIELLVWVALGSRALWGAAGGLVIGELA	302			

RESULT 13
D84180
Hypothetical protein Vmg0197c [Imported] - Halobacterium sp. NRC-1

C|Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #ext_change 02-Feb-2001
C|Accession: D84180
R|Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Jaek
; Leitbauer, B.; Keller, K.; Cruz, R.; Danson, M.T.; Hough, D.W.; Maddocks, D.G.; J
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A|Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.
A|Title: Genome sequence of *Halobacterium* species NRC-1.
A|Reference number: A84160; MUID:20504483; PMID:11016950
A|Accession: D84180
A|Status: preliminary
A|Molecule type: DNA
A|Residues: 1-342 <STO>
A|Cross-references: GB:AE004437; NID:g10579644; P10N:AA018808.1; GSPDB:GN00138
C|Genetics
I|Gene: VNG0197C

Best Lo

Query Match	6.1%;	Score 101.5;	DB 2;	Length 342;
Best Local Similarity	19.6%;	Pred. No. 0.65;		
Matches	63;	Conservative	43;	Mismatches 117;
				Indels 99;
				Gaps 14

QY 22 VELLIIIVETLALASFL--LIPGRGSRMWLVRVLISFIAEIVAVFSLEMEV 78
 Db 41 IPIVVV-----AGALWGMIVLQGA--LMIWRLIDITVSRAILVFACTSPA 88
 QY 79 GTV-----NNTSYKASAAARVARVGLVLEGIN-----ITLTGTPV 117
 Db 89 NNITPGQAGSPVAALLTTVSDADYETL-----LAIAGADALFVPSIGALTGTNA 143
 QY 118 HOLNEIYDNEFT-----WRKENYAERYANLLEKGL----- 150
 Db 144 YAIWTVSRRLRSVVGIVGAFVLVYGLAIVAGMRKRTTQALVAVRATPRKRLAVLER 203
 QY 151 ----PDPVLIAEKFITS-----SECG.YHQTHLAGHASTLWVAFCEFWLISNVLS 199
 Db 204 VEVPSERIERIACIESEFVASVERLADRTGVAKYIGLA--TALLIQCAIIMWVVF--IALG 259
 QY 200 TAPLXYGLALITTGAFALFGVF--ALASTSYVPLCPRLGSSALLTQYGA-- 249
 Db 260 SPTPIYVPLFVAVGTAKIGIPTPGAGGGIESYHVALLLTALSTAAAPLVAAAVVHRIQG 319

QY 250 FVYTLATGYLCLFLGAVVSLQ 271
 Db 320 FVLTITVG-----GGSIAVLR 335

RESULT 14

AG0648

nitrite extrusion protein (nitrite facilitator) [imported] - *Salmonella enterica* subsp.C/Species: *Salmonella enterica* subsp. *enterica* serovar TyphA/Note: this species has also been called *Salmonella typhi*

C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C/Accession: AG0648

R/Parikh, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, T.; Comerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, W.; Skelton, J.; Stevens, K.; A/TITLE: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov

A/Reference number: AB0502; MUID:21534947; PMID:11677608

A/Accession: AG0648

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-465 <PAR>

A/Cross-references: GB:AL513382; PIDN:CMD08370.1; PID:G16502415; GSPDB:GN00176

C/Genetics:

A/Gene: STY1287

C/Superfamily: nitrate transport protein mark

Query Match 6.1%; Score 101.5; DB 2; Length 465;
 Best Local Similarity 27.7%; Pred. No. 0.93;
 Matches 33; Conservative 18; Mismatches 45; Indels 23; Gaps 5;

QY 160 KFTPSSEPCGLYHQYHAGHYASATLWV-----AFCEWLL-SNVLLSTPAPLYGLALL 211
 Db 17 EWRPEDPAFWQGRH---IVASRNLIWISVPCLLAFVCWMLFSAAVAVLFPKGRV----- 68
 QY 212 TTGAPALFGVAFALASISVPLCPRLGSSALTTQYGAFFWTLATGYL---CLPFGAV 267
 Db 69 ----FTTDQLFMTALPVSIGALIRVPSFMVPLFGGRMTAFSTGILIVPCWMLGRV 123

RESULT 15

C83422

nitrate transporter PA1783 [imported] - *Pseudomonas aeruginosa* (strain PA01)C/Species: *Pseudomonas aeruginosa*

C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C/Accession: C83422

R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapid, K.; Lim, N.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A/TITLE: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho

A/Reference number: AB2950; MUID:20437337; PMID:10984043

A/Accession: C83422

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-403 <STO>

A/Cross-references: GB:AE004604; GB:AE004091; NID:G9947762; PIDN:AAG05172.1; GSPDB:GN001

A/Experimental source: strain PA01

C/Genetics:

A/Gene: nasa; PA1783

Query Match 6.1%; Score 101; DB 2; Length 403;
 Best Local Similarity 23.4%; Pred. No. 0.87;
 Matches 69; Conservative 41; Mismatches 91; Indels 94; Gaps 16;

QY 21 SVPLIVILVFLAASFLILIPGIRGSRMFLVVLSTLFIG---AEIYVHFSAEW 76
 Db 78 SAGTLAQVIVIAHAAAMLL---GVNSYAC-----ALLGLFLGAGASFAVALPLASQW 129
 QY 77 F-----VGVNTNTSYKAFSARVTAARGL--LVGLEGINITLT-----G 114
 Db 130 YPPOHGKAMGIAGAGSGIVLAALPAPGIAAAGMNNVFGFALLPLVLTLLVFAVARN 189

QY 115 TPVQIANETI-DY-----NEQFTWELKENYAAEYANALEKLPDPVLYIAEKTTPSSPCGL 169
 Db 190 AFQRPAPAMGDYLLKALGDRDSWFMFFYSVTFGEFL--GL-----ASTLPGR 235
 QY 170 YH-QYHL-----AGHYASATLWVAFCEWLLSNVLLSTPAPLYGGLALLTTGAPALFGVFL 224
 Db 236 FHDQYGLNPVTAAGYVTAACVFAG-----SLMRPLGALADRIIGIRSLMWYTL 284
 QY 225 ASISSVPLCPRLGSSALTTQYGAFFWTLATGYLCLFL-----GGAVVSL 270
 Db 285 AAI-----CIAVG-----FHLPSMAALGLFVAMSLGAGNGAVFQL 323

Search completed: February 23, 2004, 19:20:31
 Job time : 48 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2004, 18:17:55 ; Search time 48 Seconds

(without alignments)
347.134 Million cell updates/sec

Title: US-09-936-456-2

Perfect score: 1666
Sequence: 1 MTLMNGVLPFYPPQPRRAAGF.....GDLHKQALPDLKCTITNL 320

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	286.5	17.2	397	1 YKQ3 CAEEL	P34298 caenorhabd
2	111.5	6.7	381	1 YKQ3 CAEEL	P34298 caenorhabd
3	104	6.2	652	1 ACSSA_SALTI	Q821R0 salmonella
4	103.5	6.2	456	1 GUDP_PSEPU	P42205 pseudomonas
5	103	6.2	652	1 ACSSA_SALTY	Q82Kf6 salmonella
6	102.5	6.2	380	1 CYB_FANCA	P16674 rana catesb
7	102.5	6.2	381	1 CYB_HETTR	P34869 heterodoncu
8	101	6.1	1074	1 EMBQ_MYCSM	Q50393 mycobacteri
9	100.5	6.0	288	1 CLN8_MOUSE	P34866 caracharini
10	100.5	6.0	381	1 CYB_CARPL	Q79413 scyllorhinu
11	100.5	6.0	381	1 CYB_SCYCA	P42602 escherichia
12	100.5	6.0	414	1 YGLT_ECOLI	Q94593 rana nigrom
13	99.5	6.0	380	1 CYB_FANNI	P12129 oryza sativ
14	99	5.9	734	1 NUSC_ORYSA	P18935 diosophila
15	98.5	5.9	378	1 CYB_DROME	P18935 diosophila
16	98.5	5.9	379	1 CYB_CAPMR	P34867 caracharini
17	98.5	5.9	381	1 CYB_CARPO	P10903 escherichia
18	98	5.9	463	1 NARK_ECOLI	Q32880 poa pratens
19	98	5.9	702	1 NUSC_POAEP	Q32880 poa pratens
20	97.5	5.9	380	1 CYB_RANDY	Q94593 rana nigrom
21	97.5	5.9	380	1 CYB_RANDY	Q94593 rana nigrom
22	97.5	5.9	432	1 PPRF_BACCL	P41006 bacillus ca
23	97	5.8	652	1 ACSSA_ECO57	Q821R0 salmonella
24	96.5	5.8	378	1 CYB_COLMO	Q94593 diosophila
25	96.5	5.8	379	1 CYB_IGUIG	Q94593 diosophila
26	96.5	5.8	381	1 CYB_SPHLE	P34874 sphyrna lew
27	96.5	5.8	383	1 CYB_CAIOR	Q94593 diosophila
28	96.5	5.8	384	1 CYB_HYSAP	Q94593 diosophila
29	96.5	5.8	471	1 NORK_CAUCR	P58183 caulobacter
30	96.5	5.8	702	1 NUSC_SORBI	Q33066 chlamydomon
31	96	5.8	381	1 CYB_CHLRE	P23662 chlamydomon
32	95.5	5.7	326	1 NUIH_ASTPE	P23662 chlamydomon
33	95.5	5.7	381	1 CYB_SPHTT	P34876 sphyrna tib

34	95.5	5.7	381	1 CYB_SPHTV	P34875 sphyrna tib
35	95.5	5.7	149	1 ABAT_BOVIN	Q29449 bos taurus
36	95	5.7	359	1 LUCB_SYNY3	Q08871 synecocyst
37	95	5.7	353	1 CYB_ARATH	P42792 arabidopsis
38	95	5.7	419	1 CYB_RHOVI	P81378 rhodospirillum rubrum
39	95	5.7	739	1 NUSC_WHEAT	Q95446 triticum ae
40	94.5	5.7	378	1 CYB_ANOQU	P33501 anopheles g
41	94.5	5.7	378	1 CYB_MAMPR	P92658 mammutus p
42	94.5	5.7	380	1 CYB_RANAM	Q94593 rana amuren
43	94.5	5.7	381	1 CYB_PRIGL	P34873 priocetes gl
44	94.5	5.7	357	1 CYB_ORYSA	P14833 oryza sativ
45	94.5	5.7	530	1 UD17_HUMAN	Q9haw7 homo sapien

ALIGNMENTS

RESULT 1	ID	YKQ3 CAEEL	STANDARD	PRT	397 AA
AC	P34298	YKQ3 CAEEL	STANDARD	PRT	397 AA
DT	01-FEB-1994 (Rel. 28, Created)				
DT	28-FEB-2003 (Rel. 41, Last sequence update)				
DT	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	Hypothetical protein C06E1.3 in chromosome III.				
GN	C06E1.3				
OS	Caenorhabditis elegans.				
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;				
CC	Rhabditidae; Perideltidae; Caenorhabditis.				
CX	NCBI_Taxid=6239;				
RY	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=Bristol N2;				
RX	MEDLINE=94150718; PubMed=7906398;				
RA	Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,				
RA	Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,				
RA	Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,				
RA	Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,				
RA	Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,				
RA	Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,				
RA	Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,				
RA	Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,				
RA	Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,				
RA	Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,				
RA	Woldman P.				
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.				
RT	elegans."				
RL	Nature 368:32-38 (1994).				
RN	[2]				
RP	REVISIONS:				
RA	Waterson R.				
RL	Submitted (APR-2002) to the EMBL/GenBank/DDI databases.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL, L16559; AAA27934.2; -				
DR	WormPep; C06E1.3; CE30481.				
KW	Hypothetical protein; Transmembrane.				
FT	TRANSMEM 26 46				
FT	TRANSMEM 57 77				
FT	TRANSMEM 191 211				
FT	TRANSMEM 218 238				
FT	TRANSMEM 261 281				
FT	TRANSMEM 397 AA; 43801 MW; 3E6A11E2A8264A56 CRC64;				
SQ	SEQUENCE				
Query Match	17.2%; Score 286.5; DB 1; Length 397;				
Best Local Similarity	29.4%; Pred. No. 7,5e-16;				

Matches 89; Conservative 47; Mismatches 138; Indels 29; Gaps 9;

QY 4 KNGVLFFYQPPHRA--GFSVPLIVLIVFLAASFLILPGLGSHRMWMLRVLLSL 61
 DB 3 WFGGNSPSDYPHAAIPNNMAHAFVFSFLPLIAYIILDPGA-RKVVTYTYVLM 61
 QY 62 FIGAEIVAHFSAEWEVGTWNTSYKASARVAVGLVLEGINITLTGPPVQJN 121
 DB 62 AVGGALIASLIYPCWASGSMITVCFRSHNERIIAKIGVEIGLQKVVTLKFERLLSN 121
 QY 122 ETID-----NEQFTWRLKENYAEYANALEKGLPPVLYAEKFT---PSSPCGLY 170
 DB 122 DVLPQSDMELLYNNEGDISGSSMEALHGGLEGLPPLMISVLEYSINQDSFDWG-- 179
 QY 171 HOYHLAGHYASATLWVAFCEFWLISNVL-LSTPAPLYGGIALITGAFALFGVAFALASISS 229
 DB 180 RHYRVAGHYTHAIIWFAFCWGLSVLMLFLPHNV--KSLIATGISCILACIVYLLISP 237
 QY 230 VPICPLRLSSA-----LTYGGAAPWVTLATGVLCLFLGAGVVSQIYRPSALRTL 282
 DB 238 ---CELRIAFTGENFERVDLTATCFSPCYLIFALIGLCVLCGLGIGCEHWRITLSTFL 294
 QY 283 DQS 285
 DB 295 DAS 297

RESULT 2

QYB_SQUAC STANDARD; PRT; 381 AA.

AC Q9Z42;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Cytochrome b.
 GN MTCYB OR COB OR CYTB.
 OS Squalus acanthias (Spiny dogfish).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
 OC Elasmobranchii; Squala; Squalidae; Squalidae; Squalus.
 OX NCBI_TaxID=7797;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9091711; PubMed=9873084;
 RA Rasmussen A.S., Arason U.;
 RT "Phylogenetic studies of complete mitochondrial DNA molecules place
 RT cartilaginous fishes within the tree of bony fishes.";
 RL J. Mol. Evol. 48:118-123(1999).
 CC -1- FUNCTION: Component of the ubiquinol-cytochrome c reductase
 CC complex (complex III or cytochrome b-c1 complex), which is a
 CC respiratory chain that generates an electrochemical potential
 CC coupled to ATP synthesis (By similarity).
 CC -1- COFACTOR: Binds two heme groups non-covalently. Heme 1 (or BL or
 CC b562) is low-potential and absorbs at about 562, and heme 2 (or BH
 CC or b566) is high-potential and absorbs at about 566 (By
 CC similarity).
 CC -1- SUBUNIT: The main subunits of complex b-c1 are: cytochrome b,
 CC cytochrome c1 and the Rieske protein (By similarity).
 CC -1- SIMILARITY: Belongs to the cytochrome b family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Y18134; CAA77061.1; -
 CC PIR: T11546; T11546.
 CC InterPro: IPR005798; Cytb_b6_C.
 CC InterPro: IPR005797; Cytb_b6_N.
 CC Pfam: PF00032; cytochrome_b_c1_1.

DR Pfam; PF00032; Cytochrome b N; 1.
 DR PROSITE; PS00192; CYTOCHROME B HEME; 1.
 DR PROSITE; PS00193; CYTOCHROME B COO; 1.
 KW Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
 KW Heme.
 FT METAL 84 84 IRON 1 (HEME B562 AXIAL LIGAND).
 FT METAL 98 98 IRON 2 (HEME B566 AXIAL LIGAND).
 FT METAL 183 183 IRON 1 (HEME B562 AXIAL LIGAND).
 FT METAL 197 197 IRON 2 (HEME B566 AXIAL LIGAND).
 SQ SEQUENCE 381 AA; 43260 MW; BD251559304A7730 CRC64;

Query Match 6.7%; Score 111.5; DB 1; Length 381;
 Best Local Similarity 25.8%; Pred. No. 0.091;

Matches 70; Conservative 35; Mismatches 103; Indels 63; Gaps 14;

QY 27 VILVFLALASFL-LILPGLGSHRM-----FWLVRVLLST-----FIGAEIVAHFSAEW 76
 DB 118 VILVFLWATAVGVYLP-----WQMSWGAFTVITNLSAFPIGDMIV-----QW 164
 QY 77 FVGTVN---TNTSYKASARVAVRAGL-LVGLGGINITLTGPPVQJNLTID----- 125
 DB 165 IWGGSIDNATLTRFAHFLPLIIVGLTLHLFLHETGSNNPM-GLNSDMKISFHP 223
 QY 126 ---VNEQFTWRLKENYAEYANALEKGLPPVLYAEKFTPSSPCGLYHOYHLAGHYASA 182
 DB 224 YESYKDLGCFILMITLLALLFLPNLGD-----ANFTIPANP--LVTPPHIKPM--- 273
 QY 183 TLWVAFCEFWLISNVLSTPAPLYGGIALITGAFALFGVAFALASISSVPLCPLRLSSAL 242
 DB 274 -----YELFAVAIIRSIIPNLTGGVLA-----LPSIFILMIPMLHTSKORSNIIFP 320
 QY 243 TTYGGAAPWVTLATGVLCLFLGAGVVSQIY 273
 DB 321 WTQF--LFWTLVANAIIITWIGQPVQEPFI 349

RESULT 3

QYB_SALTI STANDARD; PRT; 652 AA.

AC O8Z1E0;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Acetyl-coenzyme A synthetase (EC 6.2.1.1) (Acetate--CoA ligase) (Acy]-
 DE activating enzyme).
 GN ACS OR STY4473 OR T4181.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebanina M.,
 RA Baker S., Basnam D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jorgels K.,
 RA Krogsh A., Larsen T.S., Leather S., Moulé S., O'Gaora P., Parry C.,
 RA Quail M.A., Rutherford K., Simmonds M., Skellern J., Stevens K.,
 RA Whitehead S., Barrett B.G.;
 RT "Complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhi CT18.";
 RL Nature 413:848-852 (2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Ty2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
 RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
 RT and CT18.";

```

RL J. Bacteriol. 185:2330-2337 (2003).
CC -I- FUNCTION: Enables the cell to use acetate during aerobic growth to
CC generate energy via the TCA cycle, and biosynthetic compounds via
CC the glyoxylate shunt. Acetylates chex, the response regulator
CC involved in flagellar movement and chemotaxis (By similarity).
CC -I- CATALYTIC ACTIVITY: ATP + acetate + CoA = AMP + diphosphate +
CC acetyl-CoA.
CC -I- PTM: Acetylated. Deacetylation by the SIR2-homolog deacetylase
CC activates the enzyme (By similarity).
CC -I- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL627282; CAD09259.1; -
CC EMBL; AE016848; AAC071645.1; -
CC HAMAP; MF_01123; -; 1.
CC InterPro; IPR000873; AMP-bind.
CC Pfam; PF00501; AMP-binding_1.
CC PRINTS; PR00154; AMPBINDING.
CC PROSITE; PS00455; AMP_BINDING; 1.
CC Ligase; Acetylation; Complete proteome.
CC ACT SITE 517 517 BY SIMILARITY.
CC MOD RES 609 609 ACETYLATION (BY SIMILARITY).
CC SEQUENCE 652 AA; 72163 MW; 1586F56B2B1695DA CRC64;

Query Match 6.2%; Score 104; DB 1; Length 652;
Best Local Similarity 22.0%; Pred. No. 0.67;
Matches 63; Conservative 39; Mismatches 97; Indels 88; Gaps 15;

QY 65 AEIVAVH-----PSAEWFGVNTNNTSYKAFSAATYAVGLVG-----LE 106
DB 133 ARIGAVSHVITFGSPPEAIARITIDSSRLV-----ITAEQVAGSIGILKKNVDALK 207
QY 107 GINTTLTGTPV--HQNLNETIDYNE--QFTWRLKENVAEYANLEKGLP-----D 152
DB 208 NPNVTSVEHVIVLKRTGNDIDMOGRDLNWR-----DLIEKASPEHQEAMNAED 257
QY 153 PVLVLAKEFTSPSPGLYHQYHLGH--YASATLWAF-----CFWLSNVLLSTPAEL 204
DB 258 PLFLYTSGSTGKPKGVLT--TGGYLVYAATTKRYVFDYHPGDIVYCTADV----- 307
QY 205 YGGLALLTTGAFALGFVFALASISSVPLCPRLGSSALLTQYGAAFVTLATGVLCLFG 264
DB 308 ---GWTGHSYLVG-----PLACGATLTMPE-GVPMNPTPAR--MCQVVD 347
QY 265 GAVVSLQYVRPSALRTLLDQSAKCSQSRGSGPIL--GDPHLKQA 308
DB 348 KQVAVILYTAFTAIRALMAEGDKAIEGTRSLRLTIGSVGFINPEA 394

RESULT 4
GNDP PSEPU STANDARD; PRT; 456 AA.
AC P42205;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable glucarate transporter (D-glucarate permease).
GN GNDP.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OC NCBI_TaxId=303;
RP (1)
RN SEQUENCE FROM N.A.
RC STRAIN=PP3;

```

```

RA Burlingame R.P., Lauer G.D., Platz J.G., Rudd E.A., Ally A.,
RA Ally D., Backman K.C.;
RT "Nucleotide sequence of genes for glucarate dehydratase and
RT 5-keto-4-deoxyglucarate dehydratase from Pseudomonas putida pp3."
RL Submitted (JUN-1992) to the EMBL/GenBank/DBD databases.
CC -I- FUNCTION: UPTAKE OF D-GLUCARATE.
CC -I- SUBCELLULAR LOCATION: Integral membrane protein.
CC -I- SIMILARITY: Belongs to the phthalate permease family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M69160; AAA25867.1; -
CC PIR; S27616; S27616.
CC InterPro; IPR004744; Dgal transporter.
CC InterPro; IPR007114; MFS.
CC InterPro; IPR005828; Sub transporter.
CC Pfam; PF00083; sugar tr_1.
CC TIGRfam; TIGR00893; 2A0114; 1.
CC PROSITE; PS00850; MFS; 1.
CC Transmembrane; Transport.
CC FT TRANSMEM 11 31 POTENTIAL.
CC FT TRANSMEM 42 62 POTENTIAL.
CC FT TRANSMEM 92 112 POTENTIAL.
CC FT TRANSMEM 246 266 POTENTIAL.
CC FT TRANSMEM 280 300 POTENTIAL.
CC FT TRANSMEM 317 337 POTENTIAL.
CC FT TRANSMEM 342 362 POTENTIAL.
CC FT TRANSMEM 381 401 POTENTIAL.
CC FT TRANSMEM 408 428 POTENTIAL.
CC SEQUENCE 456 AA; 49779 MW; B5B80B6608C581D CRC64;

Query Match 6.2%; Score 103.5; DB 1; Length 456;
Best Local Similarity 22.8%; Pred. No. 0.49;
Matches 80; Conservative 35; Mismatches 111; Indels 125; Gaps 18;

QY 19 GFSVPLILVIL---VFILAAASFLLILPGRSHSWFLVRLSLFGAEIVAVHFA 74
DB 101 GISTAVVILFLRMVGLAARPSF-----PQ-----NATIV-----A 132
QY 75 EWEVGVNTNNTSYKAFSAAR--VTARVGLVGLGINTTLT-----GTPVHQLNETIDY 126
DB 133 SWP-PTKERGTAASAFENSAQYFAPRAVRAIDGIDRLHLRLAARVHRHGRPGHCYLAHVD 191
QY 127 NEQFTWRLKENVAEYANLEKGLP-----PVL-YLAKEFTSPSPGLYHQYH 174
DB 192 GDLAERSPPAGYALAEVAVSSPHGGLVDLESDKKDGKGGPYKWDIYRQLLIRMMNGIY--- 247
QY 175 LAGHYASATLWAFCFWLSNVLLSTPAELY--GGLALLTTGARA-----L 218
DB 248 -LGGFCINALTYFFLWTF-----PYLVQERGMTILKGIITASPAICGFLGVLT 296
QY 219 FGVF-----ALASISSVPL-CPRLGSSALLTQYGAAFVTLATGVLCLF-- 262
DB 297 GGVTSDTLRRGNLSVARTPIVCGVLSMSNTICNYVDADMMVVCFFALAFPGKAIGA 356
QY 263 LGGAVVSLQYVRPSALRTLLDQSAKCSQSRGSGPIL--GDPHLKQA 308
DB 357 LGMAVVS-----DTPKQIAGISGLFNTFNGNLSSISPTIIG 394

RESULT 5
ACSA_SALTY STANDARD; PRT; 652 AA.
AC Q82KF6;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

```

DE Acetyl-coenzyme A synthetase (EC 6.2.1.1) (Acetate--CoA ligase) (Acy1-activating enzyme).

GN ACS OR STM4275.

OS Salmonella typhimurium.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Salmonella.

NCBI_TaxID=602;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=LT2 / SGSC1412 / ATCC 700720;

RX MEDLINE=21534948; PubMed=11677609;

RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D., Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Milvanev E., Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K.;

RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2."

RL Nature 413:852-856(2001).

RN [2]

RP ACETYLATION BY SIR2 HOMOLOG, AND MASS SPECTROMETRY.

RC STRAIN=LT2 / SGSC1412 / ATCC 700720;

RX MEDLINE=22382096; PubMed=12493915;

RA Starai V.J., Celic I., Cole R.N., Boeke J.D., Escalante-Semerena J.C.;

RT "Sir2-dependent activation of acetyl-CoA synthetase by deacetylation of active lysine."

RL Science 298:2390-2392(2002).

RN [3]

RP X-RAY CRYSTALLOGRAPHY (1.75 ANGSTROMS) AND OF ACETYLATED PROTEIN (2.3 ANGSTROMS).

RX MEDLINE=22515404; PubMed=12627952;

RA Gulick A.M., Starai V.J., Horswill A.R., Homick K.M., Escalante-Semerena J.C.;

RT "The 1.75 Å crystal structure of acetyl-CoA synthetase bound to adenosine-5'-propylphosphate and coenzyme A."

RL Biochemistry 42:2866-2873(2003).

CC -1- FUNCTION: Enables the cell to use acetate during aerobic growth to generate energy via the TCA cycle, and biosynthetic compounds via the glyoxylate shunt. Acetylates chey, the response regulator involved in flagellar movement and chemotaxis (By similarity).

CC -1- CATALYTIC ACTIVITY: ATP + acetate + CoA = AMP + diphosphate + acetyl-CoA.

CC -1- SUBUNIT: Monomer.

CC -1- PTM: Acetylated. Deacetylation by the SIR2-homolog deacetylase activates the enzyme.

CC -1- MASS SPECTROMETRY: MW=733.4; METHOD=MALDI; RANGE=607-612.

CC -1- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AE008900; AAL23099.1; ..

CC PDB; 1PG3; 03-JUN-03.

CC PDB; 1PG4; 03-JUN-03.

CC StryGene; SG7777; acs.

CC HAMAP; MF_01123; ..1.

CC InterPro; IPR000873; AMP-bind.

CC Pfam; PF00501; AMP-binding; 1.

CC PRINTS; PR00154; AMPBINDING.

CC PROSITE; PS00455; AMP BINDING; 1.

CC KX LIGASE; Acetylation; 3D-structure; Complete proteome.

CC ACT SITE 517

CC MOD_RES 609

CC SEQUENCE 652 AA; 72152 MW; 347209D1D3349D8 CRC64;

Query Match 6.24; Score 103; DB 1; Length 652;

Best Local Similarity 22.04; Pred. No. 0.81;

Matches 63; Conservative 39; Mismatches 97; Indels 88; Gaps 15;

QY 65 AEIVAVH-----FSAEMFVGTNTNTNRYKAFSAARVAVGLVG-----IE 106

DB 153 ARIQAVHVIQVGGSPPEAVAGRIIDSSRLV-----ITLBEGVRAGRSITLKNVDALK 207

QY 107 GINTITGTGV--HQLNETIDYNE--QFVRLKENVAAEYANLEKGLP-----D 152

DB 208 NENVTSEHVIVLKRGTSDIDMQEGRLMWR-----DLIEKASPEHQEAMNAED 257

QY 153 PVLVAKEFTPSQGLHYQHLAGH--YASATLMVAF-----CEFLSNVLTSPAPL 204

DB 258 PFLITVTSQSGKRGVLEHT--TGGIVVAITIKYVFDYHPGDIYCTADV----- 307

QY 205 YGGLALITLTGAFALFGVAFALASISSVPLCPRLGSSALTTQYGAAPVTLATGYLCLFLG 264

DB 308 ----GWTGHSGLYLG-----PLACGATLTME-GVNPMPTPAR--MCQVVD 347

QY 265 GAVVSLQYVRPSALRTLLDQSAKDCSQERGSPIL---GDPLHKQ 308

DB 348 KQVNLITVAPTALRALMAEDKAIEGTDRSSRLTSGVGEPTNPEA 394

RESULT 6

CYB_RANCA STANDARD; PRT; 380 AA.

AC P16674; Q9T6R7;

DT 01-AUG-1990 (Rel. 15, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Cytochrome b.

GN MTCYB OR COB OR CYTB.

OS Rana catesbeiana (Bull frog).

OC Mitochondrion.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Batrachia; Batrachia; Anura; Neobatrachia; Ranoidae; Rana.

OX NCBI_TaxID=8400;

RN [1]

RP SEQUENCE FROM N.A.

RA Lee J.-E., Lee H.-Y., Yang S.-Y.;

RT "Genetic relationships of mitochondrial cytochrome b gene among six Korean Rana species."

RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE OF 267-380 FROM N.A.

RX MEDLINE=88033615; PubMed=2444617;

RA Yoneyama Y.;

RT "The nucleotide sequences of the heavy and light strand replication origins of the Rana catesbeiana mitochondrial genome."

RL Nippon Ika Daigaku Zasshi 54:429-440(1997).

CC -1- FUNCTION: Component of the ubiquinol-cytochrome c reductase complex (complex III or cytochrome b-c1 complex), which is a respiratory chain that generates an electrochemical potential coupled to ATP synthesis (By similarity).

CC -1- COFACTOR: Binds two heme groups non-covalently. Heme 1 (or BL or b562) is low-potential and absorbs at about 562, and heme 2 (or BH or b566) is high-potential and absorbs at about 566 (By similarity).

CC -1- SUBUNIT: The main subunits of complex b-c1 are: cytochrome b, cytochrome c1 and the Rieske protein (By similarity).

CC -1- SIMILARITY: Belongs to the cytochrome b family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AF205089; AAF17088.1; ..

CC EMBL; D12694; BAA02191.1; ..

CC InterPro; IPR005798; Cytb_b6_c.

DR InterPro: IPR005797; Cytochrome b6 N.
 DR Pfam: PF00032; Cytochrome b6 C; 1.
 DR PROSITE: PS00193; CYTOCHROME_B_HEME; 1.
 DR PROSITE: PS00193; CYTOCHROME_B_HEME; 1.
 KW Electon transport; Mitochondrion; Respiratory chain; Transmembrane;
 Heme.
 FT METAL 84 84 IRON 1 (HEME B562 AXIAL LIGAND).
 FT METAL 98 98 IRON 2 (HEME B566 AXIAL LIGAND).
 FT METAL 183 183 IRON 1 (HEME B562 AXIAL LIGAND).
 FT METAL 197 197 IRON 2 (HEME B566 AXIAL LIGAND).
 FT CONFLICT 297 297 L -> F (IN REF. 2).
 FT CONFLICT 300 301 MV -> LI (IN REF. 2).
 FT CONFLICT 304 304 R -> L (IN REF. 2).
 FT CONFLICT 324 324 T -> I (IN REF. 2).
 FT CONFLICT 328 328 A -> T (IN REF. 2).
 FT CONFLICT 344 344 G -> V (IN REF. 2).
 FT CONFLICT 350 350 T -> I (IN REF. 2).
 FT CONFLICT 355 355 A -> T (IN REF. 2).
 FT CONFLICT 369 369 S -> T (IN REF. 2).
 FT CONFLICT 380 380 I -> V (IN REF. 2).
 SQ SEQUENCE 380 AA; 42345 MW; DB7C4A7252D135C9 CRC64;
 Query Match 6.2%; Score 102.5; DB 1; Length 380;
 Best Local Similarity 23.2%; Pred. No. 0.48; Mismatches 96; Indels 63; Gaps 14;
 Matches 61; Conservative 43; Mismatches 96; Indels 63; Gaps 14;
 QY 27 VILVFLALASFL-LILPGIRGSRW-----FVLVRYLSL-----FIGAEIVAHFSAEW 76
 DB 118 VILFLVMATAFVGYLP-----WGQMSFGATVITNLASAFYIGSDLV-----QW 164
 QY 77 FVG--TYNTNTSYKAFSAARVAVR---GLVGLGEGINTTGTPTPHQNLNID----- 125
 DB 165 IWGFSVDNATLIRFFHFLPTLPAASMHLLPHQGSNPT--GLNSVLDKXSRHP 223
 QY 126 ---YNEQFTWRLKENYAAYANALEKGLDPVLYLAKEPTSPSCGLVQYLAHYASA 182
 DB 224 YFSYKDLFGFVIMGALASLSTFAPNLGDP-----DNFTPNAP--LVTPPHIKEMW--- 273
 QY 183 TLWAFQFWLISNVLLSTPAPLYGGLALLTTGAFALFGVAFALASISVPLCPRLGSSAL 242
 DB 274 -----YFLFAVALIRSIIPKLGVALL-----LSIMVLFRMFIHTSKLRSIMPRPIAK 323
 QY 243 TTYGGAFWTLATGVLCFLGG 265
 DB 324 T-----FFVALIANTALTWIGG 341
 RESULT 7
 CYB_HETER STANDARD; PRT; 381 AA.
 AC P34869;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Cytochrome b.
 GN MTCYB OR COB OR CYTB.
 OS Heterodontus francisci (Horn shark).
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Chondrichthyes;
 OC Elasmobranchii; Galeomorphii; Heterodontidae; Heterodontiformes;
 OC Heterodontidae; Heterodontus.
 OK NCBI_TaxID=7792;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9252907; PubMed=1579163;
 RA Martin A.P., Naylor G.J.P., Palumbi S.R.;
 RT "Rates of mitochondrial DNA evolution in sharks are slow compared
 RT with mammals.";
 RL Nature 357:153-155(1992).
 CC -I- FUNCTION: Component of the ubiquinol-cytochrome c reductase
 CC complex (complex III or cytochrome b-c1 complex), which is a
 CC respiratory chain that generates an electrochemical potential
 CC coupled to ATP synthesis (By similarity).

CC -I- COFACTOR: Binds two heme groups non-covalently. Heme 1 (or BL or
 CC b562) is low-potential and absorbs at about 562, and heme 2 (or BH
 CC or b566) is high-potential and absorbs at about 566 (By
 CC similarity).
 CC -I- SUBUNIT: The main subunits of complex b-c1 are: cytochrome b,
 CC cytochrome c1 and the Rieske protein (By similarity).
 CC -I- SIMILARITY: Belongs to the cytochrome b family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L08035; AAA97499.1; -
 DR InterPro: IPR005798; Cytochrome b6 C.
 DR InterPro: IPR005797; Cytochrome b6 N.
 DR Pfam: PF00032; Cytochrome b6 C; 1.
 DR Pfam: PF00033; Cytochrome b6 N; 1.
 DR PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
 DR PROSITE: PS00193; CYTOCHROME_B_HEME; 1.
 KW Electon transport; Mitochondrion; Respiratory chain; Transmembrane;
 Heme.
 FT METAL 84 84 IRON 1 (HEME B562 AXIAL LIGAND).
 FT METAL 98 98 IRON 2 (HEME B566 AXIAL LIGAND).
 FT METAL 183 183 IRON 1 (HEME B562 AXIAL LIGAND).
 FT METAL 197 197 IRON 2 (HEME B566 AXIAL LIGAND).
 SQ SEQUENCE 381 AA; 41192 MW; 152943FFA0A7600 CRC64;
 Query Match 6.2%; Score 102.5; DB 1; Length 381;
 Best Local Similarity 23.7%; Pred. No. 0.48; Mismatches 105; Indels 69; Gaps 14;
 Matches 65; Conservative 35; Mismatches 105; Indels 69; Gaps 14;
 QY 27 VILVFLALASFL-LILPGIRGSRW-----FVLVRYLSL-----FIGAEIVAHFSAEW 76
 DB 118 VILFLVMATAFVGYLP-----WGQMSFGATVITNLASAFYIGDITV-----QW 164
 QY 77 FVGTV-TNTSYKAFSAARVAVRVLGLVLE-----GINITLTGTPTPHQ 119
 DB 165 IWGFSIDNATLIRFFHFLPL-LIALLTMLHFLFHTGSSNNPLGLNSMDKIPRHP 223
 QY 120 LNEFTIYNEQFTWRLKENYAAYANALEKGLDPVLYLAKEPTSPSCGLVQYLAHY 179
 DB 224 YFTYKQILGFTTTL-----FLGNLYFLFN-LIGDNENFTPNP--LVTPPHIKEMW 273
 QY 180 ASATLWVAFQFWLISNVLLSTPAPLYGGLALLTTGAFALFGVAFALASISVPLCPRLGS 239
 DB 274 -----YFLFAVALIRSIIPKLGVALL-----LFSIMLLVLPFHTSKQRNT 317
 QY 240 SALTQYGAFWTLATGVLCFLGAVVSLQVY 273
 DB 318 FRPLTQ-LILFWTLVANTIIITWIGQPEVEPFI 349
 RESULT 8
 EMBL_MYCSM STANDARD; PRT; 1074 AA.
 ID EMBL_MYCSM
 AC Q50393;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable arabinosyltransferase C (EC 2.4.2.-).
 GN EMBL.
 OS Mycobacterium smegmatis.
 OC Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OK NCBI_TaxID=1772;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97287037; PubMed=9142129;

RA Telenti A., Philip W., Sreevatsan S., Bernasconi C., Stockbauer K.E.,
 RA Mieses B., Musser J.M., Jacobs W.R.;
 RT "The emb operon, a gene cluster of Mycobacterium tuberculosis involved
 in resistance to ethambutol";
 RL Nat. Med. 3:567-570(1997).
 CC -1- FUNCTION: Arabinosyl transferase responsible for the
 CC polymerization of arabinose into the arabinan of arabinogalactan.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- MISCELLANEOUS: This is one of the target of the anti-tuberculosis
 CC drug ethambutol [(S,S')-2,2'-(ethylenedinitrodi-1-butanol) EMB].
 CC EMB is a first-line drug used to treat tuberculosis. EMB inhibits
 CC the transfer of arabinogalactan into the cell wall.
 CC -1- SIMILARITY: BELONGS TO THE EMB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by, and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U46844; AAC45271.1; -
 DR PIR; T45094; T45094.
 DR InterPro; IPRO07680; Arab transf.
 DR Pfam; PF04602; arab_transf_1.
 KW Transferase; Glycosyltransferase; Transmembrane; Cell wall;
 KW Antibiotic resistance.
 FT TRANSMEM 15 37 POTENTIAL.
 FT TRANSMEM 214 236 POTENTIAL.
 FT TRANSMEM 251 273 POTENTIAL.
 FT TRANSMEM 415 437 POTENTIAL.
 FT TRANSMEM 452 474 POTENTIAL.
 FT TRANSMEM 516 538 POTENTIAL.
 FT TRANSMEM 573 595 POTENTIAL.
 FT TRANSMEM 608 630 POTENTIAL.
 FT TRANSMEM 645 667 POTENTIAL.
 FT TRANSMEM 684 706 POTENTIAL.
 FT TRANSMEM 1074 AA; 115391 MW; 68DB19FE8939C301 CRC64;
 SQ SEQUENCE

Query Match 6.1%; Score 101; DB 1; Length 1074;
 Best Local Similarity 20.1%; Pred. No. 2.1;
 Matches 62; Conservative 35; Mismatches 81; Indels 130; Gaps 14;

QY 27 VILVFLALASFLILIPGIRGRSRF---WLVRVLSLFIGAIVAVHFSAEWVGTVNT 83
 DB 224 VAMTVIALGALHLDADGRKRKRPILPSRWMSMTLDGLVSAMLVMMH-----FVGANTA 278
 QY 84 NTSYKAFSAARVTVARVGLVGLGINTITLTGPVHQLNETIDYNEQFTWLKENYAAEYA 143
 DB 279 DDGY-ILTWARVSEHGMYA-----NYRWF 304
 QY 144 NALEKGLDPVLYIAEKTPSSPCGLYHO-VHLAGHYASATLVWVF-----CPWLLS 194
 DB 305 -----TPSPFGYIVDLALMHVSTASVWMEFPLILMGACWVVIS 346
 QY 195 -----NVLISTPAPLYGGL-----ALIT- 212
 DB 347 REVIPRLGAAKHSAAMWTAAGLFIAPWLPINNGLRPEPIIALGILLTWCVERGVATS 406
 QY 213 -----TGAFLF-GVPALASISS--VPLCTPLIGSSSLTQYGAARWTL-----A 255
 DB 407 RLIPVAVAILITGLTLTFSPTGTLAAVGLLVAGLPLKTIVAHVSFRG--YWALLAPLAA 464
 QY 256 TGYLCLFL 263
 DB 465 AGVTITFL 472

RESULT 9
 CLN8_MOUSE STANDARD; PRT; 288 AA.
 ID CLN8_MOUSE
 AC Q9QUR3;

DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE CLN8 protein.
 GN CLN8.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid:10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE:99438402; PubMed:10508524;
 RA Ranta S., Zhang Y., Ross B., Lonka L., Takkunen E., Messer A.,
 RA Sharp J., Wheeler R., Kusumi K., Moie S., Liu W., Soares M.B.,
 RA Bonaldo M.F., Hirvasniemi A., de la Chapelle A., Gilliam T.C.,
 RA Lehesjoki A.-E.;
 RT "The neuronal ceroid lipofuscinoses in human EPMR and mnd mutant mice
 RT are associated with mutations in CLN8.";
 RL Nat. Genet. 23:233-236(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Salivary Gland;
 RX MEDLINE:2238257; PubMed:12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,
 RA Datchenko L., Marzina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.J.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Guarnatre P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultk S.W.,
 RA Villalon D.K., Wuzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman W., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman D.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16699-16903(2002).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic
 CC reticulum and ER-Golgi intermediate compartment (ERGIC) (by
 CC similarity).
 CC -1- DISABS: Defects in CLN8 are the cause of the phenotype motor
 CC neuron degeneration (mnd). Mnd is a naturally occurring mutant It
 CC is characterized by progressive motor system degeneration. It has
 CC intracellular autofluorescent inclusions similar to those seen in
 CC human CLN8.
 CC -1- SIMILARITY: Contains 1 TLC (TRAM/LAG1/CLN8) domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by, and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF125308; AA04462.1; -
 DR EMBL; AF125307; AA04461.1; -
 DR EMBL; BC021625; AA021625.1; -
 DR MGI; MGI:1349447; Cln8.
 DR InterPro; IPRO06634; TLC.
 DR SMART; SM00724; TLC; 1.
 DR PROSITE; PS50922; TLC; 1.
 KW Transmembrane; Endoplasmic reticulum; Neuronal ceroid lipofuscinoses.
 FT TRANSMEM 26 46 POTENTIAL.
 FT TRANSMEM 71 91 POTENTIAL.
 FT TRANSMEM 103 123 POTENTIAL.
 FT TRANSMEM 131 151 POTENTIAL.

FT TRANSMEM 225 245 POTENTIAL.
 FT DOMAIN 62 262 TLIC
 FT SITE 285 288 ER-RETRVAL SIGNAL.
 SQ SEQUENCE 288 AA; 33108 MW; A11225DECDIDEAS CRC64;
 Query Match 6.0%; Score 100.5; DB 1; Length 288;
 Best Local Similarity 25.9%; Pred. No. 0.51;
 Matches 45; Conservative 17; Mismatches 75; Indels 37; Gaps 7;

QY 47 GHSRWFVLRYVLSLFIAGELVAVHFSMEFVGIVNNITSYKAFSARVATARGVLVGL 106
 DB 98 GQONCWPHITTAIFPPFENVAVLS-----NLPRFTDLFLVHHLPALGEL 147
 QY 107 GINTLTGTPVHQLNET---IDYNEQFT---WRLKENYAAEVANALEKGLPDPVLYAEK 160
 DB 148 GSAINLRAG--HYLMTLTLEMSPTFCISMWL-----LKAQMSDLPFKANQ 194
 QY 161 FPPSSPCGLYHOYHLAGHYASATLWVAFCFW--LLSNVLSTPAPVYGGALIT 212
 DB 195 WL-----MIHFHCRMITLYHMMWVCCHWDALTSILHPHMLFLFGLALIT 242

RESULT 10

CYB_CARPL

ID_CYB_CARPL STANDARD; PRT; 381 AA.

AC P34866;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Cytochrome b.
 GN MTCYB OR COB OR CYTB.
 OS Carcharhinus plumbeus (Sandbar shark).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
 OC Elasmobranchii; Galeomorphi; Galeoidea; Carcharhiniformes;
 OC Carcharhinidae; Carcharhinus.
 OX NCBI_TaxID=7808;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92252807; PubMed=1579163;
 RA Martin A.P., Naylor G.J.P., Palumbi S.R.;
 RT "Rates of mitochondrial DNA evolution in sharks are slow compared with mammals";
 RL Nature 357:153-155(1992).
 CC -!- FUNCTION: Component of the ubiquinol-cytochrome c reductase complex (complex III or cytochrome b-c1 complex), which is a respiratory chain that generates an electrochemical potential coupled to ATP synthesis (By similarity).
 CC -!- COFACTOR: Binds two heme groups non-covalently. Heme 1 (or BL or b562) is low-potential and absorbs at about 562, and heme 2 (or BH or b566) is high-potential and absorbs at about 566 (By similarity).
 CC -!- SUBUNIT: The main subunits of complex b-c1 are: cytochrome b, cytochrome c1 and the Rieske protein (By similarity).
 CC -!- SIMILARITY: Belongs to the cytochrome b family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L08032; AAA1703.1; -;
 CC InterPro; IPR005798; Cyb_b6_C.
 CC InterPro; IPR005797; Cyb_b6_N.
 CC Pfam; PF00032; cytochrome_b_6; 1.
 CC Pfam; PF00033; cytochrome_b_N; 1.
 CC PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
 CC PROSITE; PS00193; CYTOCHROME_B_QQ; 1.
 CC Electron transport; Mitochondrion; Respiratory chain; Transmembrane; Heme.

FT METAL 84 84 IRON 1 (HEME B562 AXIAL LIGAND).
 FT METAL 98 98 IRON 2 (HEME B566 AXIAL LIGAND).
 FT METAL 183 183 IRON 1 (HEME B562 AXIAL LIGAND).
 FT METAL 197 197 IRON 2 (HEME B566 AXIAL LIGAND).
 SQ SEQUENCE 381 AA; 43366 MW; 8A429923A8F17A11 CRC64;

Query Match 6.0%; Score 100.5; DB 1; Length 381;
 Best Local Similarity 23.4%; Pred. No. 0.7;
 Matches 64; Conservative 46; Mismatches 95; Indels 69; Gaps 15;

QY 27 VILVFLALASFL-LIIPGIRGSRW---FWLVRVLSI-----FIGAEIVAHPSAEW 76
 DB 118 VILFLMAATAFVGYVP-----WGQMSFWGATVITLISAFYIGMLV-----QM 164
 QY 77 FVGTVN-----TNTSYKFSARVATARGV-LVGLDEGINTLTGTP--VHQLNETIDYNG 129
 DB 165 IWGFSVDNATLIRFFHFHLEPLLIALTIHLLFLHETGSNNPLGINSADKISFHY 224
 QY 130 FTWR-----LKENYAAEVANALEKGLPDPVLYAEKFTPPSPGGLYHOYHLAGHYASAT 183
 DB 225 PSYDLGFFVMIFFLVAFALFMPNLLGD-----AENFIAPND--LVTPPHIKPEW---- 273
 QY 184 LWAFCEWLSNVLSTPAPVYGGALITTGARLGFVPLASISSVPLRLGSSALT 243
 DB 274 -----YFLFAVALIRSRIPNKLGGVLAU-----LPSIFLM-----LVPLHTSKQRS 315
 QY 244 TQYGA---FWTLATGVLCFLGAVVSLQYV 273
 DB 316 TIFRPMTQIFFWLLVANSILTWIGGPVQRPPI 349

RESULT 11

CYB_SCYCA

ID_CYB_SCYCA STANDARD; PRT; 381 AA.

AC 079413;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Cytochrome b.
 GN MTCYB OR COB OR CYTB.
 OS Scyliorhinus canicula (Spotted dogfish) (Spotted catshark).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
 OC Elasmobranchii; Galeomorphi; Galeoidea; Carcharhiniformes;
 OC Scyliorhinidae; Scyliorhinus.
 OX NCBI_TaxID=7830;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Muscle;
 RX MEDLINE=98393590; PubMed=9725850;
 RA Delarbre C., Spruyt N., Delmarre C., Gallut C., Barriol V., Janyier P., Landet V., Gachein G.;
 RT "The complete nucleotide sequence of the mitochondrial DNA of the dogfish, Scyliorhinus canicula";
 RL Genetics 150:331-344(1998).
 CC -!- FUNCTION: Component of the ubiquinol-cytochrome c reductase complex (complex III or cytochrome b-c1 complex), which is a respiratory chain that generates an electrochemical potential coupled to ATP synthesis (By similarity).
 CC -!- COFACTOR: Binds two heme groups non-covalently. Heme 1 (or BL or b562) is low-potential and absorbs at about 562, and heme 2 (or BH or b566) is high-potential and absorbs at about 566 (By similarity).
 CC -!- SUBUNIT: The main subunits of complex b-c1 are: cytochrome b, cytochrome c1 and the Rieske protein (By similarity).
 CC -!- SIMILARITY: Belongs to the cytochrome b family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

132 WRUKENYAAYANALEKGL---PDPVLYIAEKFTSSPCGLYHQYHLAHHYS----- 181

```

DB 155 WAIGLGFALRHGNETTKLVNDMGNVAFPMVKLVIRPAGIF-----GLVSTLATTG 208
QY 152 -ATLM-----VAFCFWLS---NVL-----STPAPYGGALATTGAPALPGVRA 223
DB 209 FSTLMGVYQALLVWLVGCMVLVALVYNPLVWVKIRNPFPLV-LLCIREGVAFFRRS 267
QY 224 LASIS-SYPLC-----PLRLGSALTQYGAAPWTLATGVCLFLGGAVVSLQY 272
DB 266 AANIPVNALCEKLNLDPDYTSVSLPLGATINMGAAITTVTLAANTLG-----IPV 322
QY 273 VAPSAIRTLDDQSANDC---SQERGSSPLIL 300
DB 323 DLPTAL--LLSVASLACAGAGVAGSILLILI 352

RESULT 13
CYS_RANNI
ID CYS_RANNI STANDARD; PRT; 380 AA.
AC Q9TER9;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Cytochrome b
OG Mitochondrion
OS Rana nigromaculata (Japanese pond frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Ranidae; Rana.
RN [1]
RP SEQUENCE FROM N.A.
RA Lee J.-E., Lee H.-Y., Yang S.-Y.;
RT "Genetic relationships of mitochondrial cytochrome b gene among six
RT Korean Rana species.";
RT Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
CC -I- FUNCTION: Component of the ubiquinol-cytochrome c reductase
CC complex (complex III or cytochrome b-c1 complex), which is a
CC respiratory chain that generates an electrochemical potential
CC coupled to ATP synthesis (By similarity).
CC -I- COFACTOR: Binds two heme groups non-covalently. Heme 1 (or BL or
CC b562) is low-potential and absorbs at about 562, and heme 2 (or BH
CC or b566) is high-potential and absorbs at about 566 (By
CC similarity).
CC -I- SUBUNIT: The main subunits of complex b-c1 are: cytochrome b,
CC cytochrome c1 and the Rieske protein (By similarity).
CC -I- SIMILARITY: Belongs to the cytochrome b family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF205087; AAF17086.1; -
DR InterPro; IPR005798; Cyto_b6_C.
DR InterPro; IPR005797; Cyto_b6_N.
DR Pfam; PF00032; cytochrome_b_C; 1.
DR Pfam; PF00033; cytochrome_b_N; 1.
DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR PROSITE; PS00193; CYTOCHROME_B_QQ; 1.
KW Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
Heme.
FT METAL 84 84 IRON 1 (HEME B566 AXIAL LIGAND).
FT METAL 98 98 IRON 2 (HEME B566 AXIAL LIGAND).
FT METAL 183 183 IRON 1 (HEME B562 AXIAL LIGAND).
FT METAL 197 197 IRON 2 (HEME B566 AXIAL LIGAND).
SQ SEQUENCE 380 AA; 42204 MW; 1453BD8F0754F84A CRC64;
Query Match 6.0%; Score 99.5; DB 1; Length 380;
Best Local Similarity 22.5%; Pred. No. 0.84;

```

```

Matches 61; Conservative 45; Mismatches 102; Indels 63; Gaps 14;
QY 27 VILVFLAASFL-LILPIGRHSRW----FMIVRVLSL-----FIGEIVAVHSAAW 76
DB 118 VILFLVWATAFVGYP-----WGQSEWATVITNLSAPFIDPLV-----QW 164
QY 77 FVG--TVNTNTSYAFSAARVTA--GLVLGLEGINITLTGPVHOLNETID----- 125
DB 165 IWGFSVDNSTLTFHFHFLIPPIAASMIHLPLHOGSSNPT-GINSNIDKYSFHP 223
QY 126 ---YNEQFTWRLKENIYAAYANALEKLPDPVLYLAEKTPSPSCGLYHQYHLAGYAA 182
DB 224 YFSYKDLGFVIMGALASLSTFAPNLTGDP-----DNFTPAMP--LVTPPHIKPWW--- 273
QY 183 TLWVAFCFWLSNVLSTPAPLVGGLALTGTGAPALPGVVALASISVPLCPRLGSSAL 242
DB 274 -----YFLPXYALRLRIPKLGVALL-----LSIMVLFMPITHTSKRLMFRPIAK 323
QY 243 TTYGAAPWVTLATGVCLFLGGAVVSLQYV 273
DB 324 T-----FFWALIANLAILTWIGQPVDPPI 349

RESULT 14
NUSC_ORYSA
ID NUSC_ORYSA STANDARD; PRT; 734 AA.
AC P12129;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE NAD(P)H-quinone oxidoreductase chain 5, chloroplast (EC 1.6.5.-)
DE (NAD(P)H dehydrogenase, chain 5) (NMDH-plastoquinone oxidoreductase
DE chain 5).
GN NDRF.
OS Oryza sativa (Rice).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Eriatrionidae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=cv. Nipponbare;
RX MEDLINE=89364698; PubMed=2770692;
RA Hiratsuka J., Shimada H., Whittier R., Ishibashi T., Sakamoto M.,
RA Mori M., Kondo C., Honji Y., Sun C.-R., Meng B.-Y., Li Y.-Q.,
RA Kanno A., Nishizawa Y., Hirai A., Shinzaki K., Sugita M.;
RT "The complete sequence of the rice (Oryza sativa) chloroplast genome:
RT intermolecular recombination between distinct tRNA genes accounts for
RT a major plasmid DNA inversion during the evolution of the cereals.";
RL Mol. Gen. Genet. 217:185-194(1989).
CC -I- CATALYTIC ACTIVITY: NAD(P)H + plastoquinone = NAD(P)H(+) +
CC plastoquinol.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X15901; CAA33950.1; -
DR PIR; J00286; DERZNS.
DR Gramene; P12129; -
DR InterPro; IPR003916; NADHUB_oxrds.
DR InterPro; IPR002128; Oxidored_q1_C.
DR InterPro; IPR001516; Oxidored_q1_C.
DR Pfam; PF00361; Oxidored_q1; 1.
DR Pfam; PF01010; Oxidored_q1_C; 1.
DR Pfam; PF00662; Oxidored_q1_N; 1.
DR PRINTS; PR01434; NMDHGNASBS.

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce>)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2004, 19:10:51 ; Search time 80 Seconds
(Without alignments)
1262.073 Million cell updates/sec

Title: US-09-936-456-2

Perfect score: 1666
Sequence: 1 MTLMNGVLPYQPRHAGP.....GDPHKQALPDKITTL 320

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1312	78.8	320	11 Q9D311	Q9d311 mus musculu
2	987	59.2	320	11 Q9D7U8	Q9d7u8 mus musculu
3	883	53.0	483	4 Q8N6K9	Q8n6k9 homo sapien
4	851.5	51.1	341	11 Q8V849	Q8v849 mus musculu
5	697.5	41.9	298	4 Q96M14	Q96m14 homo sapien
6	375	22.5	474	5 Q9VJQ7	Q9vjq7 drosophila
7	254.5	15.3	363	5 Q9V441	Q9v441 drosophila
8	127.5	7.7	391	16 Q34929	Q34929 bacillus su
9	121.5	7.3	467	16 Q9K638	Q9k638 bacillus ha
10	115	6.9	467	16 Q8Z128	Q8z128 salmonella
11	115	6.9	815	16 Q9ZB01	Q9zb01 streptomyce
12	114.5	6.9	446	17 Q974G1	Q974g1 sulfolobus
13	113.5	6.8	792	16 Q821V7	Q821v7 streptomyce
14	113	6.8	467	16 Q8ZK36	Q8zk36 salmonella
15	113	6.8	509	8 Q47547	Q47547 chlamydomon
16	109	6.5	495	17 Q8TQNS	Q8tqns mechanosarc

17	108.5	6.5	512	2 Q936C7	Q936c7 streptomyce
18	108	6.5	469	17 Q8TIES	Q8ties mechanosarc
19	107.5	6.5	289	16 Q8EUP6	Q8eup6 bruceella su
20	107.5	6.5	378	8 Q9TE18	Q9tel8 ceratitlis c
21	107.5	6.5	544	16 Q9R046	Q9r046 streptomyce
22	107	6.4	380	8 Q7YH12	Q7yh12 alioceca ca
23	107	6.4	395	16 Q8U0H3	Q8u0h3 lactobacill
24	106.5	6.4	445	17 Q96X94	Q96x94 sulfolobus
25	106.5	6.4	711	8 Q85FH9	Q85fh9 adiantum ca
26	106	6.4	380	8 Q7YB82	Q7yb82 chromis cya
27	105	6.4	495	17 Q9P9F4	Q9p9f4 methanosarc
28	106	6.4	573	16 Q89QMO	Q89qmo bradyrhizob
29	105.5	6.3	326	16 Q8T2U9	Q8t2u9 ralsconia s
30	105.5	6.3	372	8 Q21685	Q21685 pseudocarch
31	105.5	6.3	647	16 Q8A1T2	Q8a1t2 coxiella bu
32	104.5	6.3	349	8 Q03547	Q03547 rhytidopone
33	104.5	6.3	380	8 Q7YB23	Q7yb23 plectroglyp
34	104.5	6.3	380	8 Q7YB05	Q7yb05 pomacentrus
35	104	6.2	324	2 Q83130	Q83130 streptomyce
36	104	6.2	343	8 Q32532	Q32532 lithacne p
37	103.5	6.2	379	8 Q03514	Q03514 rhytidopone
38	103.5	6.2	375	8 Q9T9G3	Q9t9g3 pupa strilgo
39	103.5	6.2	502	16 Q82NC5	Q82nc5 streptomyce
40	103.5	6.2	740	8 Q8W8H5	Q8w8h5 psilocum nu
41	103	6.2	380	8 Q7YB87	Q7yb87 chromis atr
42	103	6.2	380	8 Q7YB75	Q7yb75 chromis vir
43	103	6.2	403	2 Q9X4K3	Q9x4k3 pseudomonas
44	102.5	6.2	288	11 Q8BNW2	Q8bnw2 mus musculu
45	102.5	6.2	322	16 Q8YV97	Q8yv97 anabana sp

ALIGNMENTS

RESULT 1

ID	Q9D311	PRELIMINARY	PRT	320 AA.
AC	Q9D311;			
DT	01-JUN-2001 (TREMURel. 17, Created)			
DT	01-OCT-2002 (TREMURel. 22, Last sequence update)			
DE	9030623N16R1K protein (RIKEN cDNA 9030623N16 gene).			
GN	9030623N16R1K.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Colon;			
RX	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai T., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,			
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,			
RA	Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,			
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,			
RA	Blake J., Boffelli D., Bojunga N., Carrini P., de Bona M.F.,			
RA	Brownstein M.C., But C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Gustineich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,			
RA	Lions P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,			
RA	Norone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,			
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilmink L.,			
RA	Wynhaer-Borie A., Yoshida K., Hasegawa Y., Kawai H., Kohseki S.,			
RA	Hayashizaki Y.,			
RT	"Functional annotation of a full-length mouse cDNA collection.";			
RL	Nature 409:685-690(2001).			
RN	[2]			
RP	SEQUENCE FROM N.A.			

RC TISSUE=Colon;
 RA Strausberg R.;
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK019569; BAB31281.1; -
 DR EMBL: BC031111; AAH31111.1; -
 DR MGD: MGI:1914061; 9030623N16R1K.
 SQ SEQUENCE 320 AA; 35427 MW; 4B7D4C81685162F7 CRC64;

Query Match 78.8%; Score 1312; DB 11; Length 320;
 Best Local Similarity 78.8%; Pred. No. 4.3e-100;
 Matches 253; Conservative 20; Mismatches 46; Indels 2; Gaps 2;

QY 1 MTLMNGVLPFYPPQPHAAAGFVPLIVILVFLALAAFLILPGIRGSRMFLVRLYS 60
 DB 1 MTAMDVLPFYPPQPHAAAGFVPLIVILVFLALAAFLILPGIRGSRMFLVRLYS 60
 QY 61 LFIGAEIVAHFSAEWFVGTNTNTSYKAFSAARVTARVGLVLEGINTLTGTPHQOL 120
 DB 61 LFIGAEIVAHFSGDFVGRVNTNTSYKAFSPRQVHGLVHGLAGVNTLTGTPHQOL 120
 QY 121 NETIDYNEQFTWRLKENYAAYANALEKGLPDPVLYAEKFTSPSCGLHYOHLAGHYA 180
 DB 121 NETIDYNERFTWRLMEDYTKYVHAFEKGLPDPVLYAEKFTSPSCGLHYOHLAGHYA 180
 QY 181 SATLWVAFCFWILSNVL--STPAPLYGGLALTTGAFALFGVAFALASISVPLCPRLGSS 240
 DB 181 SATLWVAFCFWITANLMLSPAPLYGGLALTTGAFALFGVAFALASISVPLCPRLGSSA 240
 QY 241 ALTTQYGAAFWTLATGVLCPLGAVVSLQYVRPSALRTLLDOSAKDCS-QEKGSGFLI 299
 DB 241 VLTPLYGASFWTLATGILSLIGAVVILHYTRPSALRSPFLDSVKDCSQAQKNSPLT 300
 QY 300 LGPDLKQALPDLKCIITNL 320
 DB 301 LNNPQHOLKSPDLN-ITTL 320

RESULT 2

Q9D7U8 PRELIMINARY; PRT; 320 AA.

DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DE 9030623N16R1K protein.
 GN 9030623N16R1K.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Stomach;
 RX MEDLINE=21085660; Pubmed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Komno H., Adachi U., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Komoto S., Yamana I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Giesi C., King B., Kochia H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schiraldi L.M., Straubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
 RA Blake J., Botfeill D., Bojunga N., Caminici F., de Bonaldo X.F.,
 RA Brownstein W.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Guetlich S., Hill D., Hofmann M., Hume D.A., Kaniya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima M., Mazzarelli J., Mombearts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Siorch K.-F.,
 RA Suzuki H., Toyochika K., Wang K.H., Weitz C., Whitaker C., Wilmberg L.,
 RA Wyrshak-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.,

"Functional annotation of a full-length mouse cDNA collection."

RL Nature 409:685-690 (2001).
 DR EMBL: AK008816; BAB25910.1; -
 DR MGD: MGI:1914061; 9030623N16R1K.
 SQ SEQUENCE 320 AA; 36342 MW; 9541848376DDEB2 CRC64;

Query Match 59.2%; Score 987; DB 11; Length 320;
 Best Local Similarity 63.9%; Pred. No. 2.7e-73;
 Matches 212; Conservative 23; Mismatches 73; Indels 24; Gaps 6;

QY 1 MTLMNGVLPFYPPQPHAAAGFVPLIVILVFLALAAFLILPGIRGSRMFLVRLYS 60
 DB 1 MTAMDVLPFYPPQPHAAAGFVPLIVILVFLALAAFLILPGIRGSRMFLVRLYS 60
 QY 61 LFIGAEIVAHFSAEWFVGTNTNTSYKAFSAARVTARVGLVLEGINTLTGTPHQOL 120
 DB 61 LFIGAEIVAHFSGDFVGRVNTNTSYKAFSPRQVHGLVHGLAGVNTLTGTPHQOL 120
 QY 121 NETIDYNEQFTWRLKENYAAYANALEKGLPDPVLYAEKFTSPSCGLHYOHLAGHYA 180
 DB 121 NETIDYNERFTWRLMEDYTKYVHAFEKGLPDPVLYAEKFTSPSCGLHYOHLAGHYA 180
 QY 181 SATLWVAFCFWILSNVL--STPAPLYGGLALTTGAFALFGVAFALASISVPLCPRLGSS 240
 DB 181 SATLWVAFCFWILSNVL--STPAPLYGGLALTTGAFALFGVAFALASISVPLCPRLGSS 240
 QY 230 VLTPLYGASFWTLATGILSLIGAVVILHYTRPSALRSPFLDSVKDCSQAQKNSPLT 289
 DB 230 VLTPLYGASFWTLATGILSLIGAVVILHYTRPSALRSPFLDSVKDCSQAQKNSPLT 289
 QY 290 S-QEKGSGFLI-LGDLKQALPDLKCIITNL 320
 DB 290 SQAQKNSPLT-LNNPQHOLKSPDLN-ITTL 320

RESULT 3

Q9N6K9 PRELIMINARY; PRT; 483 AA.

DT 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DE 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Similar to RIKEN cDNA 9030623N16 gene.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain, and Lung;
 RA Strausberg R.;
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC029819; AAH29819.1; -
 SQ SEQUENCE 483 AA; 53519 MW; 54BEF64528396C18 CRC64;

Query Match 53.0%; Score 883; DB 4; Length 483;
 Best Local Similarity 58.8%; Pred. No. 1.7e-64;
 Matches 173; Conservative 40; Mismatches 73; Indels 8; Gaps 3;

QY 9 PFYAPGKPFPPMDTLIAIIMFLTALATFVILPGIRGKRLFWLRRVVTSLFIGAIL 68
 DB 9 PFYAPGKPFPPMDTLIAIIMFLTALATFVILPGIRGKRLFWLRRVVTSLFIGAIL 68
 QY 69 AHHFSAEWFVGTNTNTSYKAFSAARVTARVGLVLEGINTLTGTPHQOLNETIDYNE 128
 DB 69 AHHFSAEWFVGTNTNTSYKAFSAARVTARVGLVLEGINTLTGTPHQOLNETIDYNE 128
 QY 129 QFTWRLKENYAAYANALEKGLPDPVLYAEKFTSPSCGLHYOHLAGHYAATLWVAF 188
 DB 129 QFTWRLKENYAAYANALEKGLPDPVLYAEKFTSPSCGLHYOHLAGHYAATLWVAF 188
 QY 189 CFPLNSVTLSTPAALYGGILALTTGAFALGCV--FALASISVPLCPRLGSSALTOY 246
 DB 189 CFPLNSVTLSTPAALYGGILALTTGAFALGCV--FALASISVPLCPRLGSSALTOY 246

QY 247 GAAFVTLATGVLCFLGGAVVSQYVRBALRTLLDQSAK-----DCSQERG 295

DB 248 GPAFWITLTPTGLCVLDGLAMVAHARMQPHRLKAFENQSVDEDDMLWEGPEEG 301

RESULT 4

ID	CEVE49	PRELIMINARY;	PRT;	341 AA.
AC	CEVE49;			
DT	01-MAR-2002	(TREMBLrel. 20, Created)		
DT	01-MAR-2002	(TREMBLrel. 20, Last sequence update)		
DT	01-OCT-2003	(TREMBLrel. 25, Last annotation update)		
DE	Hypothetical protein.			
GN	BC019755.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	[nbl_TaxID=10090,			
PN				
RP	SEQUENCE FROM N.A.			
RA	Strausberg R.			
RL	Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: BC019753; AAH19753.1; ..			
DR	MGD; MGI:2384861; BC019755.			
KQ	Hypothetical protein.			
QV	SEQUENCE 341 AA; 37595 MW; CC16D7AD103A7276 CRC64;			

RESULT 5

ID	OS6M14	PRELIMINARY;	PRT;	298 AA.
AC	OS6M14			
DT	01-DEC-2001	(T-EMBLrel. 19, Created)		
DT	01-DEC-2001	(T-EMBLrel. 19, last sequence update)		
DT	01-OCT-2002	(T-EMBLrel. 22, last annotation update)		
DE	Hypothetical protein FLJ52334.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Prostate;			
RA	Ireibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,			
RA	Hochu T., Hirooka S., Murakawa K., Takiguchi S., Kusano J.,			
RA	Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,			
RA	Sugiyama T., Irie K., Otsuki T., Sato H., Wakamatsu A., Ishi S.,			
RA	Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,			

RA Submitted (OCT-2001) to the EMBL/Genbank/DBD databases.
 RA
 RA Kikura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi A., Kanda K.,
 RA Matsutsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
 RA Suzuki Y., Sugeno S., Nagahara K., Matsuo Y., Nagai K., Isogai T.,
 RT "MDO human cDNA sequencing project."
 RI Submitted (OCT-2001) to the EMBL/Genbank/DBD databases.

RESULT 6

ID	PRELIMINARY;	PRT;	474 AA.
AC	09YU07; 095RJ0;		
DT	01-MAY-2000 (TrEMBLrel. 13, Created)		
DT	01-OCT-2002 (TrEMBLrel. 22, Last sequence update)		
DT	01-MAY-2003 (TrEMBLrel. 23, Last annotation update)		
DE	CG4482 protein (LD27791p).		
GN	BG:DS01219.1 OR CG4482 OR CG15268.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Prelygota;		
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephyraoidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Berkelley;		
RX	MEDLINE=20196006; PubMed=10731132;		
RA	Adams M.D., Celisnker S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Amaratides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,		
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,		
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,		
RA	Brandon R.C., Rogers Y.-H.C., Blasei R.G., Champe M., Pfeiffer B.D.,		
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,		
RA	Abail J.F., Agayrani A., An H.-U., Andrews-Pfannkoch C., Baldwin		
RA	Ballew R.W., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,		
RA	Besson K.Y., Bencos P.V., Bertman B.P., Bhandari D., Bolshakov S.,		
RA	Boukaya D., Bocchan M.R., Bouck J., Brokstein P., Brothier P.,		
RA	Burtis K.C., Busam D.A., Butler H., Cardieu E., Center A., Chandra I.,		
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,		
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,		
RA	Dosson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,		
RA	Dubin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,		

RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodet A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwan C.,
 RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Moberly C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Paclab J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Switzer S., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang Q., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 [2]
 RN SEQUENCE FROM N.A.
 RP Ceiniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Barton J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Chapple M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Drenek D., Farfan D.,
 RA Ferreira M., Frise E., Galle R.F., Gary N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howard T.J.,
 RA Ibegwan C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson K.A., Nunoo J.,
 RA Paclab J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phuanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Switzer S., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,
 RT "Sequencing of *Drosophila melanogaster* genome."
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [3]
 RN SEQUENCE FROM N.A.
 RP Miura S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tuny J.L., Bergman C., Berman B., Carlson J.W., Ceiniker S.E.,
 RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Milburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutnack F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.D., Lewis S.E.,
 RT "Annotation of *Drosophila melanogaster* genome."
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [4]
 RN SEQUENCE FROM N.A.
 RP Adams M.D., Ceiniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.,
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [5]
 RN SEQUENCE FROM N.A.
 RP PLAYBASE:
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 [6]
 RN SEQUENCE FROM N.A.
 RP STRAINBerkeley,
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nunoo J., Paclab J., Paragas V., Park S., Phuanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Ceiniker S.,
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AE003645; AA053425.2;
 DR EMBL: AY061346; AL28894.1;
 DR FLYBASE: FBgn0028528; BG:DS01219.1.
 SQ SEQUENCE 474 AA; 53348 MW; C48A5008304CDE64 CRC64;

Query Match 22.5%; Score 375; DB 5; Length 474;
 Best Local Similarity 34.4%; Pred. No. 1,3e-22;
 Matches 93; Conservative 46; Mismatches 115; Indels 16; Gaps 9;
 QY 22 VPLLIVILFLAALSPILLIPGRGSRMFLVRLISLFGIAEIVAHFSAEMFYGV 81
 Db 29 VIVAVSVLFAFVFAVFIIPPVGR-KKFTFTSTVLSLFGVILITRLGSAHVAHA 87
 QY 82 NNTTSKAFSAFARVAVRGLVGLGINTTGTGPV-HQNETIDYNEQFTWRKENYAA 140
 Db 88 TTIAPKAFSRKLPARGTGILGHVAVITLPIGMPRPDIDYERFTWEANDMSA 147
 QY 141 EYANALEKLPDPVLYLAETPSSPCGLV-HQYHLAGHVASATLWAFCFWILSVLLS 199
 Db 148 NTRHALQRLPPIITVAEYFSLGREGFSWCGYRAAGYAFASIMLWASLWMLNILL- 206
 QY 200 TAPLYGLALTITGAPFL---FGYFALASISSVPLCP--RLGSALLTQYGAAPWTL 254
 Db 207 IAVPRYGAVKALITGLLVCTTVGYHCL---LKRPLSLHSGRLERFGWCYLV 261
 QY 255 ATGVLCFLFGGAVS-LQYRPSALRTLLD 283
 Db 262 VAGIIC-FLAGVLSIIDLWPHRTSTYLE 290
 RESULT 7
 ID Q9V441 PRELIMINARY; PRT; 363 AA.
 AC Q9V441;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-OCT-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE BG:DS01219.1 protein.
 GN BG:DS01219.1 OR CG4482 OR CG15268.
 OS *Drosophila melanogaster* (Fruit fly).
 CC Eukaryote; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; *Drosophila*.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA Brandon R.C., Rogers Y.-H.C., Blazy R.G., Chapple M., Henderson S.N.,
 RA Sutton G.G., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA George R.A., Tector C., Berman B., Bhandari D., Bolshakov S.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.W., Cavley S., Danke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodet A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwan C.,
 RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Moberly C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Paclab J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Morley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*,"
 RL Science 287:2185-2195 (2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=99403001; PubMed=10471707;
 RA Ashburner M., Mitra S., Roote J., Lewis S.E., Blazee R., Davis T.,
 RA Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,
 RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
 RA Palazzolo M., Reese M.G., Spradling A., Tang G., Wan K., Whitelaw K.,
 RA Celniker S., Rubin G.M.,
 RT "An exploration of the sequence of a 2.9-Mb region of the genome of
 RT *Drosophila melanogaster*: the *Adh* region,"
 RL Genetics 153:1179-219 (1999).
 DR EMBL; AE003645; AAG22432.1; -;
 DR EMBL; AE003411; AAF4903.1; -;
 DR FLYBase; FBgn0028528; BG:DS01219.1.
 SQ SEQUENCE 363 AA; 41170 MW; 1912DA313B6F8446 CRC64;

Query Match 15.3%; Score 254.5; DB 5; Length 363;
 Best Local Similarity 35.3%; Pred. No. 8e-13;
 Matches 65; Conservative 30; Mismatches 74; Indels 15; Gaps 8;
 QY 108 INITLTGTPV-HQNETIDYNEQFTWRLKENYAAEANALEKGLPDPVLYAEKTPSSP 166
 DB 3 VNVTLTAIPICGNWTPPIDVNERETMEGANDMSANRHALQRLPPIITVAEYFSLGRE 62
 QY 167 CGLY-HQYHLAGHNASATLWAFPCFLLSVLSTPRLPYGSLALTITAPAL--FGVF 222
 DB 63 GFSWGGYPRAGYFASITLWASLWMLNMLL-IAVFRYGAMKALTGALLVCTTVGH 121
 QY 223 ALAISISVPLCPL--RLGSSALITQYGAATVTLATGVLCLFLGAVVS-LQYRPSALR 279
 DB 122 CL-----LPKRPLIHIEGGRLEHFFHPCWVIVLVGILC-FLAGVLISLIDVWHTS 175
 QY 280 TLID 283
 DB 176 TLVE 179

RESULT 8
 ID 034929 PRELIMINARY; PRT; 391 AA.
 AC 034929;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE YFKE protein.
 GN YFKE
 OS *Bacillus subtilis*.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 NC NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunet F., Ogatawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azeredo V., Bortier M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Bortiss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codan J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
 RA Denton F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Ertongon J., Fabre C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleon N.,
 RA Gim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Giesepp G., Guy B.U., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,

RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber U., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Maul C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
 RA Paro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Pressecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schwoerer R., Scofield F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorek N., Taccoti E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Takakoshi A., Tanaka T., Tempstra P., Tononi A.,
 RA Totato V., Uchiyama S., Vandemol M., Vanlier F., Vassarotti A.,
 RA Varti A., Wambolt R., Wedler E., Wedler H., Wellenzeger T.,
 RA Watters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
 RT *subtilis*,"
 RL Nature 390:249-256 (1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Kunet F., Ogatawara N., Yoshikawa H., Danchin A.,
 RA Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AC327;
 RA Sekiguchi J., Yamamoto H., Uchiyama S., Fajet A.,
 RA Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
 DR EMBL; D83967; CAB2620.1; -;
 DR EMBL; D83967; BAA23396.1; -;
 DR PIR; B69808; B69808.
 DR GO; GO:0016021; C:Integral to membrane; IEA.
 DR GO; GO:0005215; F:Transporter activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR007114; MFS.
 DR InterPro; IPR005828; Sub transporter.
 DR Pfam; PF00063; sugar_tr_1.
 DR PROSITE; PS00850; MFS_1.
 DR KW Transmembrane; Complete proteome.
 SQ SEQUENCE 391 AA; 42690 MW; F909AD12700104DD CRC64;

Query Match 7.7%; Score 127.5; DB 16; Length 391;
 Best Local Similarity 25.3%; Pred. No. 0.026;
 Matches 75; Conservative 33; Mismatches 105; Indels 83; Gaps 16;

QY 3 LNWGVL--PFYQPRHAGFVPLIV--ILVFLALASFLILPGIRGSRWVLRV 57
 DB 48 LYIGVLASPFMEAPFRKLGFK-PLIVMGSIYIILPFIWL-----QSWWVF-- 96
 QY 58 LLSFLIGAEIVAVFSAEMFVGTVNTNTSYKAFSAFVARGLVGLGINTLTGPFV 117
 DB 97 LRLFLIGDPMHIFSGITV---TSMSSKQ-NRGRNLSITGLSGL-GFAAGFPMPL 150
 QY 118 HQNETIDY-----NEQFTWRLKENYAAEANALEKGLDPPVLYAEKTPSSP----- 166
 DB 151 VKLSPLSPFIYSGISLFAWL-----VFLLNAYPEISPHXTSD 191
 QY 167 --CGLYHQYHLAGHNASATLWAFPCFLLSVLSTPRLPYGSLALTITAPALGCVFAL 224
 DB 192 NSFRFPQAMFLG-----WVAF-----MPTGYGLFLEALNGSPVYALRG 233
 QY 225 ASISVDP--LCPRLGDS-----SALTTOYGA--AFWTLATGVLCFLGAVVS 269
 DB 234 ISVDAAVAILPAPAFIGIITIQFPLIGLSDKYGRRNVLLVLTGALCFPIAGVPS 289
 RESULT 9
 Q9KEG8
 ID Q9KEG8 PRELIMINARY; PRT; 401 AA.

AC Q9KGC8; 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Hypothetical protein BH0884.
 GN BH0884.
 OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=86665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-125 / JCM 9153;
 RX MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kunara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 RT halodurans and genomic sequence comparison with Bacillus subtilis.";
 RL Nucleic Acids Res. 28:4317-4331(2000).
 DR EMBL; AP001510; BAB04603.1; -.
 DR FIR; D83760; D83760.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005215; P:transporter activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR007114; MFS.
 DR InterPro; IPR005828; Sub transporter.
 DR Pfam; PF00083; sugar tr_1.
 DR PROSITE; PSS0850; MFS_1.
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 401 AA; 44284 MW; E7386A8B1DCE016 CRC64;

Query Match 7.3%; Score 121.5; DB 16; Length 401;
 Best Local Similarity 24.4%; Pred. No. 0.084;
 Matches 73; Conservative 45; Mismatches 94; Indels 87; Gaps 19;
 QY 3 LNNGLV---PEYQPRHAGSVPVLLIVLFLALASFLILPGINGSRMFWLVLL 59
 DB 56 LVIGILLSPLEKELRRYGYK-PIITIGLAMIVS---LIVLP-LMQAFWEFV---L 106
 QY 60 SLFGAEIVAVHFSSEWFGVTVN-----TNTSYKAFSAARVTRVGLVGLSGINITLT 113
 DB 107 KMTIGADNMVFAQVMTTTSPPMKGRNISTYGA-----FGLGFLGPFM----- 155
 QY 114 GTPVHQLNETIDY-----NEQFTW-----RLKENYAAEYANALEKGLDPVLYAEKTPS 164
 DB 156 -TRLLQINFLPFIISAITSPFAMLLIIRLNREYPAQ---DVEIG----- 196
 QY 165 SPQGLVHOYHLAGHYASATLWAFPCFMLLSNVLTSTPAPLYG---GLAL---LTGGAFA 217
 DB 197 SQSGITRYKAVVKGWFMALPAPFCYGLHESLHGN-FPVYGLRSGLTVEQVSYLLPAFV 255
 QY 218 LRGVPALASISSVPLCPFL-RLGSSALTQYGAPFWTLATVGLCLPLGAV-VSLQYV 273
 DB 256 VGGI-----ITQMPILGFLSDRMGRKPL-----LWVLFPGATFPMYML 295

RESULT 10

Q82128 PRELIMINARY; PRT; 467 AA.
 ID Q82128;
 AC Q82128;
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
 DE Putative membrane protein.
 GN STY4802 OR T4498.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;

RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
 RA Baker S., Baaham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Felwell T., Hamlin N., Hague A., Hen T.T., Holroyd S., Jags K.,
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrett B.G.;
 RT "Complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhi CT18.";
 RL Nature 413:848-852(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Ty2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RA Deng W., Liu S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burland V., Kodoyianni V., Schwartz D.C., Blatter F.R.;
 RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
 RT and CT18.";
 RL J. Bacteriol. 185:2330-2337(2003).
 DR EMBL; AL627283; CAD06924.1; -.
 DR EMBL; AE016849; AA071945.1; -.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0015556; F:C4-dicarboxylate transporter activity; IEA.
 DR GO; GO:0015740; P:C4-dicarboxylate transport; IEA.
 DR InterPro; IPR004669; DCC.
 DR Pfam; PF03606; DCC_1.
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 467 AA; 49844 MW; 5CFD623B252926 CRC64;

Query Match 6.9%; Score 115; DB 16; Length 467;
 Best Local Similarity 24.2%; Pred. No. 0.35;
 Matches 63; Conservative 28; Mismatches 75; Indels 94; Gaps 11;
 QY 27 VLVFLALASFLILPGINGSRMFWLVLLSLFGAEIVAVHFSSEWFGVTVNNTS 66
 DB 264 VLLIF---AASFAMVIGVAVLGMW-----AISGVFLAALIVGL-TRMG 307
 QY 87 YKAFSA-----ARVAVGLVGL-EGI-----NITLTGTPVHQLNETIDYNEQFTWRK 135
 DB 308 EEARSTPTDARLLGLVALILGARGIVVMNGMITHILMS----- 351
 QY 136 ENYAAEIVANALEKGLDPVLYAEKTPSPSGLYHOYHLAGHYASATLWAFCEFWLISN 195
 DB 352 -----AESLVSGI-----STTIFINVTYWL--E 372
 QY 196 VLTSTPAPLYGALALTGGAFAFGVPALASISSVPLCPRLGSSALTQYGAPFWTLA 255
 DB 373 VILSTLVSSSGGLAVLTMPIMAPLADPAHVRDLY-----VTAGVSGIVNLY 421
 QY 256 TGVLCFLG-AYVSLQYVR 274
 DB 422 TPTSAVVMGGLAIRVPRV 441

RESULT 11

Q9ZB01 PRELIMINARY; PRT; 815 AA.
 ID Q9ZB01;
 AC Q9ZB01;
 DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Putative efflux protein.
 GN SC06483 OR SC9C7.19.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;
 OC Streptomyces; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Seeger K.J., Harris D.;
 RX Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 EMBL, AL393128; CAA22731.1; -.
 DR PIR: T35970; T35970.
 DR GO: GO:0016021; C:integral to membrane; IEA.
 DR GO: GO:0005220; F:etracycline:hydrogen antiporter activity; IEA.
 DR GO: GO:0005215; F:transporter activity; IEA.
 DR GO: GO:0015904; F:etracycline transport; IEA.
 DR GO: GO:0006810; P:transport; IEA.
 DR InterPro: IPR008969; CarboxypepD_reg.
 DR InterPro: IPR008454; Cna_B.
 DR InterPro: IPR004638; Efflux_EmrB.
 DR InterPro: IPR007114; MFS.
 DR InterPro: IPR005828; Sub_transporter.
 DR InterPro: IPR001958; TCR_beta.
 DR InterPro: IPR001411; TCR_beta.
 DR Pfam: PF05738; Cna_B_2.
 DR Pfam: PF00083; sugar tr; 1.
 DR PRINTS: PRO1035; TCRTETA.
 DR PRINTS: PRO1036; TCRTETA.
 DR TIGRFAms; TIGR00711; efflux_EmrB; 1.
 DR PROSITE: PS00850; MFS; 1.
 DR Transmembrane; Complete proteome.
 SQ SEQUENCE 815 AA; 84282 MW; 9AD3C72DF086DB96 CRC64;
 Query Match 6.9%; Score 115; DB 16; Length 815;
 Best Local Similarity 24.2%; Pred. No. 0.63; Indels 134; Gaps 15;
 Matches 85; Conservative 25; Mismatches 107;

QY 218 LFGVAFALASISVPLCEPLRLGSSALITQY-----GAFWTLATGVLCLFGAVVS 269
 DB 411 VMPPLVLAQNSVR--PTDGTATSNANVYRQIGSGVGAIVGTLTPAG----- 456
 QY 270 LQYRPSALRTLDDQAKDCQGRGSPPLISDPLHKAALPDKITTNL 320
 DB 457 -----RLTDALADRIPEAG-----VGLPDAEAITPOL 484
 RESULT 12
 ID 0974G1 PRELIMINARY; PRT; 446 AA.
 AC 0974G1.
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Putative transporter.
 GN ST0691.
 OS Sulfolobus tokodaii.
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
 OC Sulfolobus.
 OX NCBI_TaxID=111955;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JCM 10545 / 7;
 RX MEDLINE=21456156; PubMed=11572479;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
 RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
 RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
 RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kusuda N., Oguchi A.,
 RA Aoki K.-I., Masuda S., Yanagishi M., Nishimura M., Yamagishi A.,
 RA Oshima T., Kikuchi H.;
 RT "Complete genome sequence of an aerobic thermocacidophilic
 Crenarchaeon, Sulfolobus tokodaii strain7.";
 RL DNA Res. 8:123-140(2001).
 DR EMBL: AP000983; BAB5697.1; -.
 DR GO: GO:0016020; C:membrane; IEA.
 DR GO: GO:0005215; F:transporter activity; IEA.
 DR GO: GO:0006810; P:transport; IEA.
 DR InterPro: IPR007114; MFS.
 DR InterPro: IPR005828; Sub_transporter.
 DR InterPro: IPR005829; Sug_transporter.
 DR Pfam: PF00083; sugar tr; 1.
 DR PROSITE: PS00850; MFS; 1.
 DR PROSITE: PS00217; SUGAR TRANSPORT 2; 1.
 DR Hypothetical protein; Complete proteome.
 SQ SEQUENCE 446 AA; 48375 MW; F4495B31897E06CB CRC64;
 Query Match 6.9%; Score 114.5; DB 17; Length 446;
 Best Local Similarity 19.8%; Pred. No. 0.36;
 Matches 65; Conservative 52; Mismatches 90; Indels 121; Gaps 12;
 QY 6 GVLFPYQPPHAGFSVPLIIVFLALASFLILPGIRGSRFWLVRVLSLFGA 65
 DB 63 GILDFGRKRMGIDVTLMSICAIQAISQNVEL-----FSRLLIKGICIA 111
 QY 66 E-----IYVHFSAEWFGTV--NTNTSYKASAAVYARVGLV 103
 DB 112 DYVLSPITVAENAEAKRKGLMTTFPAVLWGAGVPAAVDQISSIFLSSVWRVVLGV 171
 QY 104 G-LEGINITLGTFRVHQLNETIDY-----NEQFTWRLKENYA--AEVANLE 147
 DB 172 GAIPLAISVIMNR--RKLTEILOFLTKVKPDEVELOKIKTNGLGLGINVDKPKTLRLK 228
 QY 148 KGLDPVLYLAKEFTPSSPGLYHGYHLAGHYASATLWYAFCEWLLSNVLSLTPAPLYG 207
 DB 229 ASLP-----PIVASYLMILYDISST----- 250
 QY 208 LALLTGAPALPGVAFALASISVPLCEPLRLGSSALITQYGAFWTLATGVLCLFL 263
 DB 251 -----FAIYGIIVAS-----NIGLPIITFTVAQFAGIPQOLLCIYIDKV 293

QY 264 -----GGAWSLQYVRPSALRTLLD 283
 DB 294 GRKILITIGYAGVALMVMYSL--LLD 319

RESULT 13

Q82LV7 PRELIMINARY; PRT; 792 AA.
 AC Q82LV7;
 DT 01-JUN-2003 (TREMBlrel. 24, Created)
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Putative efflux protein.
 GN SAV1903.
 OS Streptomyces avermitilis.
 OC Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;
 OC Streptomycetaceae; Streptomycetaceae; Streptomyces.
 OC NCBI_TaxId=33903;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=21477403; PubMed=11572948;
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
 RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
 RA Kikuchi H., Shiba T., Sakaki Y., Hatfort M.;
 RT "Genome sequence of an industrial microorganism Streptomyces
 RT avermitilis: deducing the ability of producing secondary
 RT metabolites.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=22608306; PubMed=12692562;
 RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
 RA Sakaki Y., Hatfort M., Omura S.;
 RT "Complete genome sequence and comparative analysis of the industrial
 RT microorganism Streptomyces avermitilis.";
 RL Nat. Biotechnol. 21:526-531(2003).
 DR EMBL: AP005028; BAC69614.1;
 DR GO: GO:0016021; C: integral to membrane; IEA.
 DR GO: GO:0015520; F: tetracycline/hydrogen antiporter activity; IEA.
 DR GO: GO:0005215; F: transporter activity; IEA.
 DR GO: GO:0015904; P: tetracycline transport; IEA.
 DR GO: GO:0006810; P: transport; IEA.
 DR InterPro: IPR008969; Carboxypeptid reg.
 DR InterPro: IPR004638; Efflux_EmrB.
 DR InterPro: IPR007114; MFS.
 DR InterPro: IPR005828; Sub. transporter.
 DR InterPro: IPR001411; TCR_TecB.
 DR Pfam: PF00083; sugar_tr; 1.
 DR PRINTS: PRO1036; TCRTEB.
 DR TIGRFAMs: TIGR00711; efflux_EmrB; 1.
 DR PROSITE: PS50850; MFS; 1.
 KW Complete proteome.
 SQ SEQUENCE 792 AA; 82256 MW; EEF543D0297D9EAL CRC64;

Query Match

Best Local Similarity 24.8%; Score 113.5; DB 16; Length 792;
 Matches 79; Conservative 31; Mismatches 111; Indels 105; Gaps 15;

QY 25 LVIIVFLAALASFLILIPGIRGSRFMVIVRLSLFICAGIIVV-----HFGAEWV 78
 DB 221 LDVIGALLAVASTCAVLTSKGTREYANDSRITLIGAGACVAVFLVYEHFAPBELI 280
 QY 79 GTVNTNTSYAF--SAARVAVYGLVLEGINITLTGTPVHQLNETIDYNEQFWRLKE 136
 DB 281 -----PIRLFRDSYFNTGLVGLVIG-----VALFG----- 306
 QY 137 NYAAEY-----ANALEKGLPDPVLYIAEKFPSSPCGYCHLAGHYASAT--- 183
 DB 307 --AASTPLFLQWIDGASATSGI---LML-----PMAGIGVASTIAGGLISHTGRY 354
 QY 184 -LW-----VAFCEWILSNVLLSTPAPLYG-GLALLTGGAFALRGVAFALASISSVPLCP 234

DB 355 RMNPILIGSVLAAGVMMLSLREADTPRLQYSIMWVLTGAGIGMWPVVLAVQNSVR--P 412
 QY 235 LRIGSSALTTQYGAAPFVTLATGVLCLFPGAVWSLQYVRPSALRTLLDQSAKDSQSRG 294
 DB 413 ADLGTATSSANNYRQ-----IGSSVGA-----AVFGTLFHRADLAEER- 452

QY 295 GSPILLDDPLHKQALPDLKCTITNL 320
 DB 453 -----LPTRAGARLPDPESLTPOL 471

RESULT 14

Q82K36 PRELIMINARY; PRT; 467 AA.
 AC Q82K36;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
 DE Putative arginine repressor.
 GN STM464.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OC NCBI_TaxId=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali U., Dame M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2.";
 RL Nature 413:852-856(2001).
 DR EMBL: AE008909; AAL23283.1;
 DR GO: GO:0016021; C: integral to membrane; IEA.
 DR GO: GO:0015556; F: C4-dicarboxylate transporter activity; IEA.
 DR GO: GO:0015740; P: C4-dicarboxylate transport; IEA.
 DR InterPro: IPR004669; Dcuc.
 DR Pfam: PF03606; Dcuc; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 467 AA; 49890 MW; 0650E1B5F0475FE CRC64;

Query Match 6.8%; Score 113; DB 16; Length 467;
 Best Local Similarity 23.8%; Pred. No. 0.51;
 Matches 62; Conservative 29; Mismatches 75; Indels 94; Gaps 11;

QY 27 VLIIVFLAALASFLILIPGIRGSRFMVIVRLSLFICAGIIVVHFGAEWFGVNTNTS 86
 DB 264 VLMIF--AAAPAVITGVAVLGMW-----AETSGVFLAAIIVGI-TRMG 307
 QY 87 YKAPSA-----ARVAVYGLVGL-EGI-----NITLTGTPVHQLNETIDYNEQFWRLK 135
 DB 308 EEAFTSTFLDGRDLGVALLIGARGLVVMWDGMITHTLHS----- 351
 QY 136 ENYAAEYANALEKGLPDPVLYIAEKFTSSPCGLYHQYHLAGHYSATLWVAFQWLTSLN 135
 DB 352 -----AESLVSGL-----STTIFLNTYTWL--E 372
 QY 196 VILSTPAFLPYGGLALLTGAFLFGVAFALASISSVPLCPRLIGSSALTTQYGAAPFWTLA 255
 DB 373 VILSTFLVPSGSGLATVLMPIVAPLADFAHYGRDLV-----VTAVQASAGIVNLV 421
 QY 256 TGVLCIFLGG-AVVSLOQYR 274
 DB 422 TPTSAYVMGGLAIARVPYR 441

RESULT 15

Q47547

AC 047547; PRELIMINARY; PRT; 509 AA.
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Cytochrome c oxidase subunit I (EC 1.9.3.1) (Cytochrome c oxidase polypeptide I).
 DE COX1.
 GN Chlamydomonas eugametos.
 OS Mitochondrion.
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 OC Chlamydomonadaceae; Chlamydomonas.
 NCBI_TaxID=3053;
 [1]
 RP SEQUENCE FROM N.A.
 RP MEDLINE=98145434; PubMed=9484440;
 RX Denoyan-Wright E.M., Nedelcu A.M., Lee R.W.;
 RA "complete sequence of the mitochondrial DNA of Chlamydomonas eugametos";
 RT Plant Mol. Biol. 36:285-295(1998).
 RL [2]
 RP SEQUENCE FROM N.A.
 RP Denoyan-Wright E.M., Nedelcu A.M., Lee R.W.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2 AND HEME A OF SUBUNIT 1 TO THE METALLIC CENTER FORMED BY HEME A3 AND COPPER B (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERROCYTOCHROME C + 2 H2O.
 CC -1- PATHWAY: RESPIRATORY CHAIN, TERMINAL STEP.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL INNER MEMBRANE (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY. EMBL; AF008237; AAC3942.1; -.
 DR PIR; T11043; T11043.
 DR HSSP; P98002; IAR1.
 DR GO; GO:0019866; C:inner membrane; IEA.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0005739; C:mitochondrial electron transport chain; IEA.
 DR GO; GO:0009481; P:a3-type cytochrome c oxidase; IEA.
 DR GO; GO:0009482; P:a3-type cytochrome c oxidase; IEA.
 DR GO; GO:0009483; P:a3-type cytochrome c oxidase; IEA.
 DR GO; GO:0009485; P:cbb3-type cytochrome c oxidase; IEA.
 DR GO; GO:0004129; P:cytochrome-c oxidase activity; IEA.
 DR GO; GO:0016491; P:oxidoreductase activity; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR000883; COX1.
 DR Pfam; PF00115; COX1; 1.
 DR PRINTS; PRO1165; CYCOXIDASE1.
 DR PROSITE; PS00077; COX1; 1.
 DR Copper; Electron transport; Heme; Inner membrane; Membrane; Oxidoreductase; Respiratory chain; Transmembrane; Transport; Mitochondrion.
 KW SEQUENCE 509 AA; 55847 MW; 8EAA06905EA00E12 CRC64;

Query Match 6.8%; Score 113; DB 8; Length 509;
Best Local Similarity 21.2%; Pred. No. 0.56;
Matches 66; Conservative 38; Mismatches 118; Indels 90; Gaps 14;

[illegible]

QY 116 PHQQLNETIDVNEQSTWMLKENVAAEYVNALEKGIIDPVLYLA-----EXETP 163
Db 169 GMAIIMKISQMPLEFWMSIA-----FTALIVIAEVLAAALVMLLTDRNLNTAYFCE 221
QY 164 SSPCGLY-HQYHLAGHYASATLMTWAFCEWLLSNVL-ISTPAPLYGIALP-TTGAFALFG 220
Db 222 SGDLIIYQGLFMFFGH-PEVYIILIPAGIYSHVISFSSQCKPIIGNGMGCAMGALITLG 280
QY 221 VFALAS-----ISSVPLCPRLGS--SALTQYGAAPWYT---LATGV 258
Db 281 FIYVAHMFFTVGHLDLTIAVFTSATMTIAVPTGMKIFSMIATIVGSLMLITTPMWEAVGF 340
QY 259 LCLFLGGAAYSL 270
Db 341 ICLFLTGGVTGV 352

Search completed: February 23, 2004, 19:19:33
Job time : 83 secs

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

OM nucleic - nucleic search, using sw model

10984.465 Million cell updates/sec

Sequence: 1 ctctagcgtgccgctctgcc.....aaaaaaataaaagtcgacc 498

apop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size :

Total number of hits satisfying chosen parameters: 6940544

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Listing first 45 summaries

Database :

1: gb_ba:*
2: gb_hng:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_px:*
10: gb_ro:*
11: gb_scs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_lm:*
20: em_om:*
21: em_ov:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_scs:*
28: em_un:*
29: em_vl:*
30: em_hng_hum:*
31: em_hng_inv:*
32: em_hng_other:*
33: em_hng_mus:*
34: em_hng_pln:*
35: em_hng_rod:*
36: em_hng_mam:*
37: em_hng_vrt:*
38: em_sy:*
39: em_hng_hum:*
40: em_hng_mus:*
41: em_hng_other:*

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Query Match	Length	DB	ID	Description
1	48	100.0	498	6	AX035348	AX035348 Sequence
2	45	85.3	1441	6	AX035346	AX035346 Sequence
3	32	78.7	2684	6	AX174511	AX174511 Sequence
4	32	78.7	2684	6	AX056896	AX056896 Homo sapi
5	32	78.7	156534	2	AC009700	AC009700 Homo sapi
6	32	78.7	166937	2	AC012255	AC012255 Homo sapi
7	32	78.7	171444	2	AC087790	AC087790 Homo sapi
8	32	78.7	181312	9	AC091117	AC091117 Homo sapi
9	37	75.1	1740	9	HS080624	HS080624 Homo sapi
10	217	43.6	217	9	HS189812	HS189812 Homo sapi
11	16	33.3	1923	9	BC029819	BC029819 Homo sapi
12	16	27.3	218	9	HS185411	HS185411 Homo sapi
13	26	5.2	214866	9	AC068544	AC069544 Homo sapi
14	26	5.2	254436	3	AE014827	AE014827 Plasmid
15	25	55790	8	AB025602	AB025602 Arabidops	
16	24	4.8	95824	8	U90439	U90439 Arabidopsis
17	24	4.8	115811	9	AL136377	AL136377 Human DNA
18	24	4.8	157965	2	AC011220	AC011220 Homo sapi
19	24	4.8	186184	2	AC140136	AC140136 Homo sapi
20	24	4.8	188769	2	AP002809	AP002809 Homo sapi
21	24	4.8	204467	9	AC010736	AC010736 Homo sapi
22	24	4.6	90157	9	AL355552	AL355552 Human DNA
23	23	4.6	91851	8	ATT16129	AL138659 Arabidops
24	23	4.6	125623	3	AC115529	AC115529 Dictyoste
25	23	4.6	137764	2	AC008655	AC008655 Homo sapi
26	23	4.6	153368	5	EX649231	EX649231 Zebrafish
27	23	4.6	153789	2	AC027472	AC027472 Homo sapi
28	23	4.6	154677	2	AC016196	AC016196 Homo sapi
29	23	4.6	154677	2	AC016196	AC016196 Homo sapi
30	23	4.6	157326	2	AC011024	AC011024 Homo sapi
31	23	4.6	161079	2	AC036148	AC036148 Homo sapi
32	23	4.6	182881	9	AC090980	AC090980 Homo sapi
33	23	4.6	183016	2	AC068611	AC068611 Homo sapi
34	23	4.6	192125	2	AC145790	AC145790 Silurana
35	23	4.6	193265	9	AC105314	AC105314 Homo sapi
36	23	4.6	193265	9	AC090797	AC090797 Homo sapi
37	23	4.6	194122	2	AC138778	AC138778 Homo sapi
38	23	4.6	230307	2	AC097902	AC097902 Rattus no
39	23	4.6	250029	9	AE014310	AE014310 Homo sapi
40	22	4.4	6415	6	AX251125	AX251125 Sequence
41	22	4.4	93680	9	AC005265	AC005265 Homo sapi
42	22	4.4	98387	9	AL603790	AL603790 Human DNA
43	22	4.4	108864	8	AC073391	AC073391 Oryza sat
44	22	4.4	110000	1	AP006828	AP006828 Homo sapi
45	22	4.4	110250	9	AC010104	AC010104 Homo sapi

RESULT 1
AY03E349

LOCUS	AX035348	498 bp	DNA	PAT 15-NOV-2000
DEFINITION	Sequence 3 from Patent WO0053748.		linear	

VERSION AX035348.1 GI:111910655

SOURCE Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1

TITLE	Novel compounds
JOURNAL	Patent: WO 0053748-A 3 14-SEP-2000;


```
1.1441
/organism="Homo sapiens"
/mol_type="unassigned DNA"
```

```

Query Match      78.7%   Score 352;   LB 6;   Length 2683;
Best Local Similarity 100.0%;   Pred.No. 4.2e-209;
Matches 392;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0

```

Db 2243 GGGGCTCTGTGCTCTTCTCTGAGAGGGGCGGTGTGATCTCCAGTATGTTGGGCCAGC 2284

Qy 151 GCTCTTGTGACCCCTTCTGTGACCAAGCGCCAAAGAGTGTGACGACCGAGAGAGGGGCTCA 210

Db 2283 GCTCTTGTGACCCCTTCTGTGACCAAGCGCCAAAGAGTGTGACGACCGAGAGAGGGGCTCA 2224

Qy 211 CCTCTATCTCTGGGCGGCGCCACTGTGACCAAGAGGGGCGCTCTCCAGACTTAATGATTC 270

Db 2223 CCTCTATCTCTGGGCGGCGCCACTGTGACCAAGAGGGGCGCTCTCCAGACTTAATGATTC 2164

Qy 271 ACCACTAAGCTGTGAGGGGAGCCCAATCTGTGACTCTTCCCGCTTGGGACATGCAAG 330

Db 2163 ACCACTAAGCTGTGAGGGGAGCCCAATCTGTGACTCTTCCCGCTTGGGACATGCAAG 2104

Qy 331 CCGGAGAGCAAGTGGGCGGCGCCAGAGGCTGGGCGGAGAGTCCAGAGAGGGCACTGACGCT 390

Db 2103 CCGGAGAGCAAGTGGGCGGCGCCAGAGGCTGGGCGGAGAGTCCAGAGAGGGCACTGACGCT 2044

Qy 391 GCTGGCGCGAGGCTCTGGACATCGGACGAGCAGGAGAAAGTCTCTGGGCGGATCTGTA 450

Db 2043 GCTGGCGCGAGGCTCTGGACATCGGACGAGCAGGAGAAAGTCTCTGGGCGGATCTGTA 1984

Qy 451 AATAACCTTTTCTTTCTTTGTTTTTAAAAA 482

Db 1983 AATAACCTTTTCTTTCTTTGTTTTTAAAAA 1952

RESULT 4

AK056896/c

LOCUS AK056896 2684 bp mRNA linear PRI 01-AUG-2002

DEFINITION Homo sapiens cDNA FLJ32334 f1s, clone PROST2005426.

ACCESSION

AK056896.1 GI:16552419

oligo capping; f1s (full insert sequence).

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

location/Qualifiers

1..2684

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="PROST2005426"

/tissue_type="prostate"

/clone_id="PROST2"

/note="Cloning vector: pME18SFL3"

CDS

56..952

/note="unnamed protein product"

/codon_start=1

/protein_id="BAB71304.1"

/db_xref="GI:16552420"

/translation="MATTGHTPEFYAGKPPFMDTTLASIIIMIFLTAATFVILPG

IRGKTRFLMLLVVTSFLFGAIIIGTLPVQDNTENTINVEEFTMLGENVAEEVKALE

KGLPDVYIAEKTPRSDGIVQYRAGHTSAMLVAFGLANVMVSMPLVY

GGYMLATGIRIQLALLPSPMATSLSGCPHILASVYHTHGPAFMTITLGLCTL

LEIAMVAHRHQPHRLKAFPNQSDDEDPMLEWSPBEGSLSPRISMADSPKSDIPL

SSASTKAYCKEAPHKPDPCAL"

Query Match 78.7%; Score 392; DB 9; Length 2684;

Best Local Similarity 100.0%; Pred. No. 4,2e-209;

Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 91 GGGGCTCTGTGCTCTTCTCTGAGAGGGGCGGTGTGATCTCCAGTATGTTGGGCCAGC 150

Db 2343 GGGGCTCTGTGCTCTTCTCTGAGAGGGGCGGTGTGATCTCCAGTATGTTGGGCCAGC 2284

Qy 151 GCTCTTGTGACCCCTTCTGTGACCAAGCGCCAAAGAGTGTGACGACCGAGAGAGGGGCTCA 210

Db 2283 GCTCTTGTGACCCCTTCTGTGACCAAGCGCCAAAGAGTGTGACGACCGAGAGAGGGGCTCA 2224

Qy 211 CCTCTATCTCTGGGCGGCGCCACTGTGACCAAGAGGGGCGCTCTCCAGACTTAATGATTC 270

Db 2223 CCTCTATCTCTGGGCGGCGCCACTGTGACCAAGAGGGGCGCTCTCCAGACTTAATGATTC 2164

Qy 271 ACCACTAAGCTGTGAGGGGAGCCCAATCTGTGACTCTTCCCGCTTGGGACATGCAAG 330

Db 2163 ACCACTAAGCTGTGAGGGGAGCCCAATCTGTGACTCTTCCCGCTTGGGACATGCAAG 2104

Qy 331 CCGGAGAGCAAGTGGGCGGCGCCAGAGGCTGGGCGGAGAGTCCAGAGAGGGCACTGACGCT 390

Db 2103 CCGGAGAGCAAGTGGGCGGCGCCAGAGGCTGGGCGGAGAGTCCAGAGAGGGCACTGACGCT 2044

Qy 391 GCTGGCGCGAGGCTCTGGACATCGGACGAGCAGGAGAAAGTCTCTGGGCGGATCTGTA 450

Db 2043 GCTGGCGCGAGGCTCTGGACATCGGACGAGCAGGAGAAAGTCTCTGGGCGGATCTGTA 1984

Qy 451 AATAACCTTTTCTTTCTTTGTTTTTAAAAA 482

Db 1983 AATAACCTTTTCTTTCTTTGTTTTTAAAAA 1952

RESULT 5

LOCUS AC009700

DEFINITION Homo sapiens chromosome 15 clone RP11-163P10 map 15, WORKING DRAFT

SEQUENCE 15 unordered pieces.

AC009700.4 GI:7622346

HTG; HTGS_PHASE1; HTGS_DRAFT.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 156534) DNA linear HTG 20-APR-2000

Birken, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,

Baker, J., Baldwin, J., Barna, N., Beckert, R., Bern, J., Brown, A.,

Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A.,

Cooke, P., Dearielano, K., Depayre, E., Devon, K., Dewar, K.,

Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, M., Forrest, C.,

Funk, R., Gage, D., Galagan, J., Gardy, S., Gilbert, D., Grant, G.,

Hagos, B., Haefford, A., Horton, L., Howland, J., Jones, C., Kann, L.,

Karatas, N., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P.,

Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J.,

Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,

AC012255	LOCUS	AC012255	166937 bp	DNA	linear	HTG 06-MAY-2001
	DEFINITION	Homo sapiens chromosome 15 clone RP11-109D20 map 15, WORKING DRAFT				
	ACCESSION	AC012255				
	VERSION	AC012255.4	GI:7684440			
	KEYWORDS	HTG; HTGS PHASE1; HTGS_DRAFT.				
	SOURCE	Homo sapiens (human)				
	ORGANISM	Homo sapiens				
	REFERENCE	Eukaryota; Metazoa; Chordata; Cranita; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
	AUTHORS	1 (bases 1 to 166937)				
	TITLE	Birren,B., Linton,L., Nusbaum,C. and lander,E.				
	JOURNAL	Homo sapiens chromosome 15, clone RP11-109D20				
	REFERENCE	Unpublished				
	AUTHORS	2 (bases 1 to 166937)				
		Birren,B., Linton,L., Nusbaum,C., lander,E., Allen,N., Anderson,M.,				

```

misc_feature /note="assembly_fragment"
60224..73532
/note="assembly_fragment"
misc_feature 73633..87433
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
misc_feature 87534..101708
/note="assembly_fragment"
101809..121193
/note="assembly_fragment"
misc_feature 121294..139051
/note="assembly_fragment"
139152..166937
/note="assembly_fragment"
clone_end:T7
vector_side:right"

```

ORIGIN

```

Query Match      78.7%; Score 392; DB 2; Length 166937;
Best Local Similarity 100.0%; Pred. No. 6e-209;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 GGGCTCTGCTGCTCTCTCTCGAGGCGCGGTGAGTCTTCAGATGTTCCGCCCAAC 150
Db 161083 GGGCTCTGCTGCTCTCTCTCGAGGCGCGGTGAGTCTTCAGATGTTCCGCCCAAC 161142

QY 151 GCTCTTCGACCTCTTCGACCAAGCGCCAGAGCTCAGAGAGAGAGAGAGAGAGAG 210
Db 161143 GCTCTTCGACCTCTTCGACCAAGCGCCAGAGCTCAGAGAGAGAGAGAGAGAGAG 161202

QY 211 CCTCTTATCTCTGCGCAGCCCACTGCAAGAGAGCGCGCTCTTCCAGACTTAAATGATC 270
Db 161203 CCTCTTATCTCTGCGCAGCCCACTGCAAGAGAGCGCGCTCTTCCAGACTTAAATGATC 161262

QY 271 ACACATACCTGTGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 330
Db 161263 ACACATACCTGTGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 161322

QY 331 CCGGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 390
Db 161323 CCGGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 161382

QY 391 GCTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 450
Db 161383 GCTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 161442

QY 451 AATTAACCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 482
Db 161443 AATTAACCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 161474

```

```

RESULT 7
AC087790 171444 bp DNA linear HTG 09-MAY-2001
LOCUS Homo sapiens chromosome 15 clone RP11-276K9 map 15, WORKING DRAFT
DEFINITION AC087790.13 unordered pieces.
ACCESSION AC087790.2 GI:13357344
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 171444)
AUTHORS Birren B., Linton L., Nusbaum C. and Lander E.
TITLE Homo sapiens chromosome 15, clone RP11-276K9
JOURNAL Unpublished
2 (bases 1 to 171444)
AUTHORS Birren B., Linton L., Nusbaum C., Lander E., Allen N., Anderson S.,
Barna N., Bastien V., Boguslavsky L., Boukhgalter B., Brown A.,
Camataia J., Campoliano A., Choquet Y., Colangelo M., Collins S.,
Collymore A., Cooke P., Dearellano K., Dewar K., Diaz J.S.,

```

TITLE JOURNAL COMMENT

Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karatas, A., Lacroque, K., Lamazares, R., Landers, T., Lenczky, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Margulis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPherson, R., Meldrum, J., Menus, L., Mihova, T., Menga, V., Murphy, T., Navlot, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rie, C., Rogov, P., Roman, O., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Sougnier, C., Spencer, B., Strange-Thomann, N., Stojanovic, N., Strausen, A., Subramanian, A., Talamas, J., Testaye, S., Theodores, J., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, M.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (24-JAN-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 16, 2001 this sequence version replaced g1:12408501.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Genome Center
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu

Project Information
Center project name: 112451
Center clone name: 276_K_9

Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 165906 bases at least Q40
Consensus quality: 168587 bases at least Q30
Consensus quality: 169557 bases at least Q20
Insert size: 180000; agarose-IP
Insert size: 170244; sum-of-coverage
Quality coverage: 5.7 in Q20 bases; agarose-IP
Quality coverage: 6.0 in Q20 bases; sum-of-coverage

NOTE: This is a 'working draft' sequence. It currently consists of 13 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 6091: contig of 6091 bp in length
6092 6191: gap of 100 bp
6192 7986: contig of 1795 bp in length
7987 8086: gap of 100 bp
8087 43939: contig of 35853 bp in length
43940 44039: gap of 100 bp
44040 45298: contig of 1259 bp in length
45299 45398: gap of 100 bp
45399 47595: contig of 2197 bp in length
47596 47695: gap of 100 bp
47696 52385: contig of 4690 bp in length
52386 52485: gap of 100 bp
52486 56310: contig of 3825 bp in length
56311 56411: gap of 100 bp
56411 71413: contig of 14903 bp in length
71413 71414: gap of 100 bp
71414 90993: contig of 15580 bp in length
90994 91093: gap of 100 bp
91093 111459: contig of 20366 bp in length
111460 111559: gap of 100 bp
111560 137726: contig of 26167 bp in length

* 137727 137826: gap of 100 bp
 * 137827 170053: contig of 32227 bp in length
 * 170054 170153: gap of 100 bp
 * 170154 171444: contig of 1291 bp in length.
 Location/Qualifiers

FEATURES

Source

1. 171444
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="15"
 /map="15"
 /clone="RP11-276X9"
 /clone_1id="RP11-11 Human Male BAC"
 1. .6091
 /note="assembly_fragment"
 clone end:SP6
 vector_side:left"

misc_feature
 6192..7986
 /note="assembly_fragment"
 8087..43939
 /note="assembly_fragment"
 44040..45298
 /note="assembly_fragment"
 45399..47595
 /note="assembly_fragment"
 47696..52385
 /note="assembly_fragment"
 52486..56310
 /note="assembly_fragment"
 56411..71313
 /note="assembly_fragment"
 71414..90993
 /note="assembly_fragment"
 91094..111459
 /note="assembly_fragment"
 111560..137726
 /note="assembly_fragment"
 137827..170053
 /note="assembly_fragment"
 170154..171444
 /note="assembly_fragment"
 clone end:T7
 vector_side:right"

ORIGIN

Query Match 78.7%; Score 392; DB 2; Length 171444;
 Best Local Similarity 100.0%; Pred. No. 6e-209;
 Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 GAGCGTCTGTCCTCTCTCTCGAGAGGCGGTGAGTCTCCAGTATGTTCCGCCACG 150
 Db 143780 GGGCTCCCTGCTCTCTCTCTCGAGAGGCGGTGAGTCTCCAGTATGTTCCGCCACG 143839
 QY 151 GCTCTTGGCACCCTTCTTGAGCAAAAGCGCCAGAGCTGCAAGCCAGAGAGGGGCTCA 210
 Db 143840 GCTCTTGGCACCCTTCTTGAGCAAAAGCGCCAGAGCTGCAAGCCAGAGAGGGGCTCA 143899
 QY 211 CCTTTATCTCTCGGAGCCTGCACTGCAAGAGGCGCTCTCCAGACTTAATATGATC 270
 Db 143900 CCTTTATCTCTCGGAGCCTGCACTGCAAGAGGCGCTCTCCAGACTTAATATGATC 143959
 QY 271 ACCACTAACCCTGTGAGGGGAGCCCAATCTGAGCTCTTCCCGCTTGGAGCAATCGCAGG 330
 Db 143960 ACCACTAACCCTGTGAGGGGAGCCCAATCTGAGCTCTTCCCGCTTGGAGCAATCGCAGG 144019
 QY 331 CCGGAGAGAGTCCCGCCAGAGGCGCTGAGAGAGCTCCAGAGAGGCACTGAGGCT 390
 Db 144020 CCGGAGAGAGTCCCGCCAGAGGCGCTGAGAGAGCTCCAGAGAGGCACTGAGGCT 144079
 QY 391 GCTGGCGGAGGCTCTGAGCAATCCGACAGCACCAGGAAAGTCTCTCGGGCGCATCTGTA 450
 Db 144080 GCTGGCGGAGGCTCTGAGCAATCCGACAGCACCAGGAAAGTCTCTCGGGCGCATCTGTA 144139

QY 451 AATAACCTTTTCTTTTCTTTTAAAA 482
 Db 144140 AATAACCTTTTCTTTTCTTTTAAAA 144171

RESULT 8

AC091117 181312 bp DNA linear PRI 20-FEB-2002
 LOCUS Homo sapiens chromosome 15 clone RP11-109D20 map 15q15, complete
 DEFINITION sequence.
 AC091117
 AC091117.5 GI:18767395
 HTG.
 Homo sapiens (human)
 SOURCE
 ORGANISM

REFERENCE

AC091117

1 (bases 1 to 181312)
 Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
 Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
 Pate,D. and Hood,L.
 Sequencing of human chromosome 15 DIS146-D15S117 region
 Unpublished
 2 (bases 1 to 181312)
 Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
 Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
 Pate,D. and Hood,L.
 Direct Submission
 Submitted (30-MAR-2001) Multimegabase Sequencing Center, Institute
 for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
 98105, USA

REFERENCE

AC091117

3 (bases 1 to 181312)
 Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
 Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
 Pate,D. and Hood,L.
 Direct Submission
 Submitted (13-FEB-2002) Multimegabase Sequencing Center, Institute
 for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
 98105, USA

REFERENCE

AC091117

4 (bases 1 to 181312)
 Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
 Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
 Pate,D. and Hood,L.
 Direct Submission
 Submitted (14-FEB-2002) Multimegabase Sequencing Center, Institute
 for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
 98105, USA

REFERENCE

AC091117

5 (bases 1 to 181312)
 Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
 Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
 Pate,D. and Hood,L.
 Direct Submission
 Submitted (20-FEB-2002) Multimegabase Sequencing Center, Institute
 for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
 98105, USA
 On Feb 20, 2002 this sequence version replaced gi:18653547.

----- Genome Center
 Center: Multimegabase Sequencing Center
 Center code: UMSC
 Web site: http://chroma.mbl.washington.edu/msg_www
 Contact: leetowens@systemsbiology.org
 ----- Summary Statistics
 Sequencing vector: pUC18; 108752
 Chemistry: Dye-terminator Big Dye; 90% of reads
 Chemistry: Dye-primer Big Dye; 10% of reads
 Assembly program: Phrap; version 0.990399

Note: Data from overlapping clone AC090888 [Drafting Center: WIBRI] was added for finishing. RP11-109D20 spans bases 1-174236 of this sequence. Bases 174237-179373 derive from clone RP11-276X9, AC087790 [Drafting Center: WIBRI] to establish the overlap with the next BAC in our tiling path. Bases 179373-181312 derive from

362 GGAGAGCTCCAGGAGGCACTGAGCGCTGCTGGCCGAGGCGCTCGGACATCCGAGGCA 421

Db 1605 GGAGAGCTCCAGGAGGACCTGAGGCTGCTGGCGCGAGGCTCGACATCCGAGGCA 1664

Qy 422 CCAGGAAAGTCTCCCTGGGGCGATCTGTAAATAAAGCTTTTCTTTGTTTTTAAAA 481

Db 1665 CCAGGAAAGTCTCCCTGGGGCGATCTGTAAATAAAGCTTTTCTTTGTTTTTAAAA 1724

Qy 482 AAAAA 486

Db 1725 AAAAA 1729

RESULT 10
HS183H12F/c 217 bp DNA linear PRI 18-OCT-1995
LOCUS HS183H12F
DEFINITION H.sapiens Cpg island DNA genomic MseI fragment, clone 183h12,
forward read cpg183h12.fta.

ACCESSION 257552
VERSION 257552.1 GI:1028793
KEYWORDS Cpg island; genomic MseI fragment.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1
AUTHORS Cross, S.H., Charlton, J.A., Nan, X. and Bird, A.P.
TITLE Purification of Cpg islands using a methylated DNA binding column
JOURNAL Nat. Genet. 6 (3), 236-244 (1994)
MEDLINE 94282070
PUBMED 8012384

REFERENCE 2 (bases 1 to 217)
AUTHORS Dodsworth, S.J., Huckle, E., Wilkinson, P. and Micklem, G.
TITLE Direct Submission
JOURNAL Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, England. E-mail contact: humquerry@sanger.ac.uk
Vector: pGEM-5Zf(-)
COMMENT Clones are available from the UK MRC Human Genome Mapping Project
Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. see URL:
http://www.hgmp.mrc.ac.uk/ for details
or contact: DioneId@hgmp.mrc.ac.uk.
Location/Qualifiers
1. 217
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="183h12"
/sex="male"
/tissue_type="blood"
/clone_id="CGI-1"
/dev_stage="adult"

FEATURES
source

ORIGIN

Query Match 43.6%; Score 217; DB 9; Length 217;
Best Local Similarity 100.0%; Pred. No. 1,6e-110;
Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 262 AAATGATACACCACTAACTGTGAGGGGAGCCCAATCTGACTCTTCCCGGCTTGGA 321

Db 217 AAATGATACACCACTAACTGTGAGGGGAGCCCAATCTGACTCTTCCCGGCTTGGA 158

Qy 332 CATGGCAGCGCGGAGAGAGTGCCTCCAGGCTTGCGCCAGAGAGTCTCCAGGAAGGGA 361

Db 157 CATGCAGCGCGGAGAGAGTGCCTCCAGGCTTGCGCCAGAGAGTCTCCAGGAAGGGA 98

Qy 382 CTGAGCGCTGCTGGCGGAGGCTCTGACATCCGACAGGACCAAGGAAAGTCTCTGGGG 441

Db 97 CTGAGCGCTGCTGGCGGAGGCTCTGACATCCGACAGGACCAAGGAAAGTCTCTGGGG 38

Qy 442 CGATCTGTAATAAACCCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 478

Db 37 CGATCTGTAATAAACCCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1

RESULT 11

BC029819/c 1923 bp mRNA linear PRI 06-OCT-2003
LOCUS BC029819
DEFINITION Homo sapiens hypothetical protein FLJ22334, mRNA (cDNA clone
MGC:35310 IMAGE:5177619), complete cds.
ACCESSION BC029819
VERSION BC029819.1 GI:20987583
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1
AUTHORS Strausberg, R.D., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buettow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Schuetz, T.E., Brownstein, M.J., Usdin, T.B., Tostitvski, S.,
Carninci, P., Prange, C., Rana, S.S., Loquellano, N.A., Peters, G.J.,
Abramsen, R.D., Mullany, S.J., Bosak, S.A., Moysan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, R.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Faney, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Scherchenko, Y.,
Boutifard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalka, U., Smalls, D.E.,
Schnerk, A., Schein, J.E., Jones, S.J. and Marra, M.A.
TITLE human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE 22388257
PUBMED 12477932

REFERENCE 2 (bases 1 to 1923)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (06-MAY-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgabs@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunnar, F.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louie, H.,
Kwis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.

FEATURES
source

1. 1923
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:35310 IMAGE:5177619"
/tissue_type="Brain, Lung, Testis, adult, pooled whole"
/clone_id="NH_MGC_115"
/lab_host="DH10B"
/note="Vector: pCMV-Sport6"
1. 1923

gene

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LINL at: http://image.llnl.gov
Series: IRAC plate: 51 Row: 0 Column: 3
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 21389332.
Location/Qualifiers


```
/gene="FLJ32334"
/db_xref="locusID:90527"
401..1852
/codon_start=1
/product="FLJ32334 protein"
/protein_id="AAH29819.1"
/db_xref="GI:20987584"
/db_xref="locusID:90527"
/translation="MATHGTFPPYAGKPPFMDTLASIMFLPALTFIVLPG
IRGKRLFWLIRVYSLFGAIIIVANSSMSVSGYSTNTSYKAPSSMSIDILG
VGGGNTITLTCTPYQINERTINNEETRLGNNYAEYAKLEKLPDPVYTLAK
FTFRSPCGIYQYRLAGHTSMWVAFICWLANVMSPLVYGGIMLALATGIFOL
IALTFEPMATSLTSPCLGASVLTHTHGAFTITTTGLCYLGLAAVAHRCQP
RLSKAFENQSDVDEPMLEWSPBEGILSPRYSMADSPKSDIPLISASSTKYRPR
RLSLVADVAGLAPALALPGALIAQMARLLPGLRCPKXKESRLGPHSPWRGP
EGCEERMAHETGDSRPLRGRTGRLNMGSKERRACGVRAMLPRLVNSGLKRPSC
DLPKCMVDRDARAFPHLEPTPCTSHRTPLI"

ORIGIN
Query Match 33.3%; Score 166; DB 9; Length 1923;
Best Local Similarity 100.0%; Pred. No. 1.1e-81;
Matches 166; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 CTGTAGGGGGGAGCCCATCTGACCTCTCCCGCTTGAGGACATCGAGGCGGAGAGC 339
DB 1557 CTGTAGGGGGGAGCCCATCTGACCTCTCCCGCTTGAGGACATCGAGGCGGAGAGC 1498
QY 340 AGTCCCGCCAGGCGCTGAGGCGCAGAGAGCTCCAGAGAGGCACTGAGCGCTGCGCGC 399
DB 1497 AGTCCCGCCAGGCGCTGAGGCGCAGAGAGCTCCAGAGAGGCACTGAGCGCTGCGCGC 1438
QY 400 AGGCTTCGACATCCCGCAGGACCAAGGAAAGTCTCCGCGGCGAT 445
DB 1437 AGGCTTCGACATCCCGCAGGACCAAGGAAAGTCTCCGCGGCGAT 1392

RESULT 12
HS183A12R 218 bp DNA linear PRI 19-OCT-1995
LOCUS H.sapiens Cpg island DNA genomic MseI fragment, clone 183a12,
reverse read cpg183a12.rta.
Z59954
Z59954.1 GI:1031867
CPG island; genomic MseI fragment.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.
Purification of Cpg islands using a methylated DNA binding column
Nat. Genet. 6 (3), 236-244 (1994)
JOURNAL 94282070
MEDLINE 8012384
PUBMED 8012384
COMMENT 2 (bases 1 to 218)
Dodsworth,S.J., Huckle,E., Wilkinson,P. and Micklem,G.
Direct Submission
Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, England. E-mail contact: humquerry@sanger.ac.uk
Vector: pGEM-5Zf(-)
Clones are available from the UK MRC Human Genome Mapping Project
Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:
http://www.hgmp.mrc.ac.uk/ for details
or contact: biocel@hgmp.mrc.ac.uk.
Location/Qualifiers
1..218
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="183a12"
/sex="male"
/tissue_type="blood"
/clone_lib="CGI-1"

FEATURES
Source
```

```
/dev_stage="adult"

ORIGIN
Query Match 27.3%; Score 136; DB 9; Length 218;
Best Local Similarity 100.0%; Pred. No. 7.6e-65;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 CGGAGAGAGTGGCCCGCAGGCGCTGAGGCGCAGAGAGCTCCAGAGAGGCACTGAGCGCTG 391
DB 72 CGGAGAGAGTGGCCCGCAGGCGCTGAGGCGCAGAGAGCTCCAGAGAGGCACTGAGCGCTG 131
QY 332 CTGGGCGCAGGCGCTCGACATCCGACAGCCAGGAAAGTCTCCGCGGCGATCTGTGA 451
DB 132 CTGGGCGCAGGCGCTCGACATCCGACAGCCAGGAAAGTCTCCGCGGCGATCTGTGA 191
QY 452 ATAAACCTTTTCT 467
DB 192 ATAAACCTTTTCT 207

RESULT 13
AC069544 214866 bp DNA linear PRI 25-APR-2002
LOCUS Homo sapiens chromosome 10 clone RP11-398C13, complete sequence.
AC069544
AC069544.9 GI:20303530
HTG.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 214866)
Smith,D.R.
Genome Therapeutics Corporation Sequencing Center: Human Genome
Sequence Data.
JOURNAL Unpublished
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
3 (bases 1 to 214866)
Smith,D.R.
REFERENCE Direct Submission
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2001) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
4 (bases 1 to 214866)
Smith,D.R.
REFERENCE Direct Submission
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
On Apr 25, 2002 this sequence version replaced gi:17975233.
COMMENT Location/Qualifiers
1..214866
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-398C13"
/clone_lib="RP11-398C13"

ORIGIN
Query Match 5.2%; Score 26; DB 9; Length 214866;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 464 TTCTTTTGTGTTTTTAAAAAATAA 489
DB 200470 TTCTTTTGTGTTTTTAAAAAATAA 200495

RESULT 14
AE014827
```

repeat_region	complement(1057, .1106)
	/rpt_type=tandem
	/rpt_unit="(A)n"
repeat_region	complement(1112, .1146)
	/note="AT_rich"
	/rpt_type=tandem
repeat_region	complement(1204, .1362)
	/note="AT_rich"
	/rpt_type=tandem
repeat_region	1382, .1405
	/rpt_type=tandem
	/rpt_unit="(A)n"
repeat_region	complement(1406, .1428)
	/note="AT_rich"
	/rpt_type=tandem
repeat_region	complement(1447, .1544)
	/note="AT_rich"
	/rpt_type=tandem
repeat_region	complement(1545, .1568)
	/rpt_type=tandem
	/rpt_unit="(TA)n"
repeat_region	complement(1644, .1690)
	/note="AT_rich"
	/rpt_type=tandem
gene	<1703, .>2211
	/locus_tag="PF14_0645"
	join(<1703, .1751,1852, .>2211)
mRNA	/locus_tag="PF14_0645"
	join(1703, .1751,1852, .1964,2125, .2211)
CDS	/locus_tag="PF14_0645"
	/codon_start=1
	/product="hypothetical protein"
	/protein_id="AA037258.1"
	/db_xref="GI:23497720"
	/translation="MENETKQPYFPAINGCESVSEPFKCNLTLPDQDLIKSS
	SQDQCPKQKXKCTDCKLKQKTPIMFLTEKENK"
	complement(1764, .1846)
repeat_region	/note="AT_rich"
	/rpt_type=tandem
repeat_region	2044, .2092
	/rpt_type=tandem
repeat_region	complement(2048, .2092)
	/rpt_type=tandem
	/rpt_unit="(TA)n"
repeat_region	complement(2097, .2123)
	/note="AT_rich"
	/rpt_type=tandem
repeat_region	complement(2239, .2366)
	/note="AT_rich"
	/rpt_type=tandem
repeat_region	complement(2367, .2396)
	/rpt_type=tandem
	/rpt_unit="(A)n"
repeat_region	complement(2398, .2418)
	/rpt_type=tandem
	/rpt_unit="(TA)n"
repeat_region	complement(2469, .2764)
	/note="AT_rich"
	/rpt_type=tandem
repeat_region	2822, .2865
	/rpt_type=tandem
repeat_region	complement(2822, .2865)
	/rpt_type=tandem
	/rpt_unit="(TA)n"
repeat_region	complement(2904, .2927)
	/note="AT_rich"
	/rpt_type=tandem
repeat_region	complement(2950, .2974)
	/note="AT_rich"
	/rpt_type=tandem
repeat_region	complement(3003, .3059)
	/note="AT_rich"
	/rpt_type=tandem

```

repeat_region complement(3093..3152)
                /note="AT rich"
                /rpt_type=tandem
gene            <3157..>3530
                /locus_tag="PF14_0646"
                /join(<3157..3281,3389..3530)
mRNA           /locus_tag="PF14_0646"
                /join(3157..3281,3389..3530)
CDS            /locus_tag="PF14_0646"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AA037259.1"
                /db_xref="GI:23497721"
                /translation="MTAKRDPNDYKICMOSHAKDKAKCEKLYKAIINSHIS
                RECLPYTEDLOQCFHSPRLSPCKEIMDKTKCQSDVNLITS"

repeat_region /rpt_type=tandem
                complement(3315..3366)
                /rpt_type=tandem
                /rpt_unit="(TA)n"
                /note="AT rich"
repeat_region complement(3533..3596)
                /note="AT rich"
                /rpt_type=tandem
                complement(3597..3638)
                /rpt_type=tandem
                /rpt_unit="(A)n"
                /note="AT rich"
repeat_region /rpt_type=tandem
                complement(3713..3789)
                /note="AT rich"
                /rpt_type=tandem
                complement(3859..3895)
                /note="AT rich"
                /rpt_type=tandem
                complement(3926..3961)
                /rpt_type=tandem
                complement(3926..3957)
                /rpt_type=tandem
                /rpt_unit="(TA)n"
                complement(3997..4137)
                /rpt_type=tandem
                /rpt_unit="(TA)n"
                /note="AT rich"
repeat_region /rpt_type=tandem
                complement(4286..4320)
                /rpt_type=tandem
                /rpt_unit="(TAAA)n"
                complement(4335..4398)
                /note="AT rich"
                /rpt_type=tandem
                complement(4529..4555)
                /rpt_type=tandem
                /rpt_unit="(A)n"

Query Match      5.2%; Score 26; DB 3; Length 254436;
Best Local Similarity 100.0%; Pred No. 0.015; 0; Indels 0; Gaps 0;
Matches 26; Conservative 0; Mismatches 0;

Qy      466 CTTTGTGTTTTTAAAAAATAAAA 491
Db      60242 CTTTGTGTTTTTAAAAAATAAAA 60267

RESULT 15
AB025602/c  AB025602  55790 bp  DNA  linear  PLN 27-DEC-2000
LOCUS      Arabidopsis thaliana genomic DNA, chromosome 5, BAC clone:FL4A1.
DEFINITION AB025602 BA000015
ACCESSION  AB025602.1 GI:4589408
KEYWORDS
SOURCE      Arabidopsis thaliana (thale cress)
ORGANISM    Arabidopsis thaliana

```

```

REFERENCE
AUTHORS      Kaneko,T., Katoh,T., Asamizu,E., Sato,S., Nakamura,Y., Kotani,H.
TITLE        Structural analysis of Arabidopsis thaliana chromosome 5. XI
JOURNAL      Unpublished
AUTHORS      2 (bases 1 to 55790)
Nakamura,Y.
DIRECT SUBMISSION
Submitted (02-APR-1999) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research, 1532-3, Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamu@kazusa.or.jp,
Tel:81-438-52-3935, Fax:81-438-52-3934)
Address for correspondence: kao@kazusa.or.jp
For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/agg_graph.cgi?c=FL4A1
Genes with similarity to proteins in the databases are described in
'product' or 'note' qualifiers. Genes that have no significant
protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Grail
(Informatics Group, Oak Ridge National Laboratory,
http://compbio.ornl.gov/Grail-1.3//),
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
SplicePredictor (Volker Brendel, Stanford University,
http://gremlin.zozi.iastate.edu/cgi-bin/sp.cgi).
Genes encoding tRNAs are predicted by tRNAscan-SE
(Sean Eddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNAscan-SE//).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is M1K22 and the 3' clone is MEB13.

FEATURES
source
1..55790
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone_1fb="FL4A1"
/clone_1fb="TIGR BAC"
complement(345..538)
/product="nucleolar protein-like"
/note="CDS is reported in Acc# AB005236
gene_id:M1K22.22"
number=2
/evidence=not experimental
complement(650..1037)
/product="nucleolar protein-like"
/note="CDS is reported in Acc# AB005236
gene_id:M1K22.22"
number=1
/evidence=not experimental
complement(2095..2228)
/note="gb|AAD23664.1
gene_id:FL4A1.2
similar to unknown protein"
pseudo
/codon_start=1
/evidence=not experimental
complement(3022..3227,3496..4530)
/note="contains similarity to unknown protein
gb|AAD38227.1
gene_id:FL4A1.3"
pseudo
/codon_start=1
/evidence=not experimental
join(8265..8758,8827..9079,10490..10842,10919..11470)
/note="gene_id:FL4A1.4
pir||T01797

```

```

strong similarity to unknown protein"
/pseudo
/codon_start=1
/evidence=not experimental
join(14755..15729,16136..16766)
/note="contains similarity to cytochrome P450
gene_id:F14A1.5"
/pseudo
/codon_start=1
/evidence=not experimental
17455..18278
/note="contains similarity to Athila ORF 1
gene_id:F14A1.6"
/pseudo
/codon_start=1
/evidence=not experimental
22069..22536
/note="gene_id:F14A1.7"
/codon_start=1
/evidence=not experimental
/product="cytochrome P450-like"
/protein_id="P4507454.1"
/db_xref="GI:8865524"
/translation="MKKLTKEAPRIHVAFPNDPMSTTVVVDGFIPIGSKLTI
SRMIGRNPVWDKPLKFDPERHLSNNTCVLEKPDINTITGAEKRGCMADIGSAM
TYMLVARIIOGFTWSPVSESKIDISVSKDLMAXPLNAVVTFRIDPVYPA"
24745..25395
/note="contains similarity to Athila retroelement ORF1
protein
gene_id:F14A1.8"
/pseudo
/codon_start=1
/evidence=not experimental
join(29275..29414,29688..30532,30645..30785,30870..31208)
/note="gb|AAD29066.1
gene_id:F14A1.9
similar to unknown protein"
/pseudo
/codon_start=1
/evidence=not experimental
complement(join(32433..32517,32622..32684,32833..33045,
33139..33248,33712..33756,33846..33905,34042..34091,
34183..34267,34610..34669,34811..34919,35405..35448,
35606..35692,35775..35850,36080..36744,37204..37923,
38243..38449,38897..39093,39661..40059,40141..40393))
/note="gene_id:F14A1.10
unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="FAA97455.1"
/db_xref="GI:8865525"
/translation="MSERKRGKPKNCISIKHFDVDAASKPHQIAVIOASGLTARSIRS
BSAAPVYDGDCKPTFAELSSIDSILRLRLISNDSDVAGSDVYMPKVALYMP
PSVEYVSVFSLRCGAFPLDPSMREVRVLSLSSNSLSIACGLSSVSHMLVE
RNVCPVLIFSMDEKLSVETGSSSFVWCKERQKFCYLMYTSGSTKRGVCGTEOG
LINFELWMOELPYVGEORPAFTSVGFIDHIOBFLAIISSPALVIPEFTLKEMMI
SIIDPLEEYSISRLAVSMIRATLPTLORGHNNKLOSLKLVVSGEPFPLSLMS
LHSLPERCPNLNGSTVEGDDCTYFDCSRLPRLITKEEISVPIGKSTSNCKVILG
DEKPYRBEICVSGCLCSQGYMSSISSEGYVLKNNSLCNHLLTNDGSQLYRRGDY
GRQLSSGDLITGRDRITVLRNGKMLBEITTELENPDAVAVLLSRDETELASL
KAFVVLNKNSSDGLIFSLRNWGGKLPVMI PNHFVLEKLPITSSGKVDYELAR
LKCPITTAQDMQONGNSLONIKKAVCDALVKEVSDDDDFALGDSLLAAHLISH
SLGIDMRLIYORSPSRLIYLSEKSKLRDMQONTTQKLDHRIEQNGNLVSRV
PLHSGVTGPTPSLCEKNNSPKRLKIDYEKSPKMKENKLMDSGFSOIOCAFRC
MKVSPSCSNEEAREYMSLEI PRNOMVMOEIMKVHMSCVDAAPLVILKDSKTYL
FTGSHSRKFSCTDAKSGMWETLEIRIGSAMVGDPSQVYICGKXKLYEDPST
GSLCKRQACGEITKQPVVDTSQLIWCGSHDITVLAIDRSQCYKACCGSIFAS
PAIDEGHSLVASTSGVIAVSTKDSPEFTLWLFELBAPITGSLCTIPSTQWDEQV
IAMPSPGIIIRYRTGPIFAGPCMSHVLPQVAVCCRNCGVSLPEBSGLWEDNI
GDPITASAYIDENLHFSHELASDRLVATVCCSSGRVHVARPSILSRSDSHSKVE
ITRMELOADPSSPVMIGRIIFVGCDDYVHCLSLSCRTTITFEKREIMISDIKIF
LEIYEIETGKITIIEKHKNGNTNL"
join(41916..41972,42875..42975,43101..43682)

```

```

/note="gb|AAD15520.1
gene_id:F14A1.11
similar to unknown protein"
/pseudo
/codon_start=1
/evidence=not experimental
complement(join(49029..49804,49883..50017,50099..51775,
51848..52158,52220..52875,52933..53719))
/note="contains similarity to retroelement pol polyprotein
gene_id:F14A1.12"
/pseudo
/codon_start=1
/evidence=not experimental
complement(54463..55327)
/note="gene_id:F14A1.13
pir|T05807
similar to unknown protein"
/pseudo
/codon_start=1
/evidence=not experimental

```

```

ORIGIN
Query Match 5.0%; Score 25; DB 8; Length 55790;
Best Local Similarity 100.0%; Pred.No. 0.048;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 459 TTTTTCCTTGTGTTTTTAAAAA 483
DB 20789 TTTTTCCTTGTGTTTTTAAAAA 20765

```

Search completed: February 24, 2004, 01:57:08
Job time : 1969.03 secs

CDS

Query Match	100.0%;	Score 498;	DB 3;	Length 498;
Best Local Similarity	100.0%;	Pred. No. 3.6e-238;		
Matches 498; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY	Db	QY	Db
1 CTCCTACGCGGCGCTCTGCGCCGCTCCCGGCTTAGGCTCTCCGCGCTACCACTCAGAACG 60	1 CTCCTACGCGGCGCTCTGCGCCGCTCCCGGCTTAGGCTCTCCGCGCTACCACTCAGAACG 60	1 CTCCTACGCGGCGCTCTGCGCCGCTCCCGGCTTAGGCTCTCCGCGCTACCACTCAGAACG 60	1 CTCCTACGCGGCGCTCTGCGCCGCTCCCGGCTTAGGCTCTCCGCGCTACCACTCAGAACG 60
QY	Db	QY	Db
61 AAGCAGCGCCTTCTGGGTCAAGCTGGCAACCGAGGCTCTGTAGTGCCTCTTCTCGAGGGGCC 120	61 AAGCAGCGCCTTCTGGGTCAAGCTGGCAACCGAGGCTCTGTAGTGCCTCTTCTCGAGGGGCC 120	61 AAGCAGCGCCTTCTGGGTCAAGCTGGCAACCGAGGCTCTGTAGTGCCTCTTCTCGAGGGGCC 120	61 AAGCAGCGCCTTCTGGGTCAAGCTGGCAACCGAGGCTCTGTAGTGCCTCTTCTCGAGGGGCC 120
QY	Db	QY	Db
121 GTGGTGAAGTCTCCAAGTAGTGTGGGCCCAAGCGCTCTTGGCACCCTTCTGAGACCAAGGCC 180	121 GTGGTGAAGTCTCCAAGTAGTGTGGGCCCAAGCGCTCTTGGCACCCTTCTGAGACCAAGGCC 180	121 GTGGTGAAGTCTCCAAGTAGTGTGGGCCCAAGCGCTCTTGGCACCCTTCTGAGACCAAGGCC 180	121 GTGGTGAAGTCTCCAAGTAGTGTGGGCCCAAGCGCTCTTGGCACCCTTCTGAGACCAAGGCC 180
QY	Db	QY	Db
181 AAGAGCTGACAGCCAGAGAGAGGGGGCTCACTCTTATCTCTGGGGACCCACTGTCACAAG 240	181 AAGAGCTGACAGCCAGAGAGAGGGGGCTCACTCTTATCTCTGGGGACCCACTGTCACAAG 240	181 AAGAGCTGACAGCCAGAGAGAGGGGGCTCACTCTTATCTCTGGGGACCCACTGTCACAAG 240	181 AAGAGCTGACAGCCAGAGAGAGGGGGCTCACTCTTATCTCTGGGGACCCACTGTCACAAG 240

QY	24	CAGCGCGCTCTCCCACTTAATAATATACCACTAACCTGTGAGGGGACCCAACTCG	3 08
Dp	241	CAGCGCGCTCTCCCACTTAATAATATACCACTAACCTGTGAGGGGACCCAACTCG	3 08
QY	301	GATCTCTCCCGCGCTTTGGGACATGGCGAGGCGGGGAACATGTCGCCGACAGCTGGAGC	3 68
Dp	301	GATCTCTCCCGCGCTTTGGGACATGGCGAGGCGGGGAACATGTCGCCGACAGCTGGAGC	3 68
QY	361	AGAGAGAGCTCCAGAGAGGGGACATGAGAGCGCTGTGGCGGAGAGCTCCGAGACATCCGAGGC	4 28
Dp	361	AGAGAGAGCTCCAGAGAGGGGACATGAGAGCGCTGTGGCGGAGAGCTCCGAGACATCCGAGGC	4 28
QY	421	ACCAAGGAAAGTCTCTCTGGGGCGATCTGTAAATTAACCTTTTTCCTTTGTGTTTTTAA	4 88
Dp	421	ACCAAGGAAAGTCTCTCTGGGGCGATCTGTAAATTAACCTTTTTCCTTTGTGTTTTTAA	4 88
QY	481	AAAAAATTTAAAGTGCAC	4 98
Dp	481	AAAAAATTTAAAGTGCAC	4 98

RESULT 2	
AAA94623	
ID	AAA94623 standard; DNA; 1441 BP.
XX	
AC	AAA94623;
XX	
DT	11-JAN-2001 (first entry)
XX	
DE	Human CASB618 coding sequence.
XX	
KW	Epitope: human; CASB618; chromosome 15; tumour; vaccine; cancer; ovarian.
KW	colon; autoimmune disease; HLA_A0201, ss.
XX	
OS	Homo sapiens.

```

PH Key Location/Qualifiers
FT CDS 259..1221
FT /*tag= a
FT /product= "Human CASB618"
TX
TX WO20053748-A2.
TX
TX
TX
TX 14-SEP-2000.
TX
TX
TX 09-MAR-2000; 2000WO-EP002048.
TX
TX 11-MAR-1999; 99GB-00005607.
TX 01-SEP-1999; 99GB-00020590.
TX
TX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
TX
TX
TX Bruck CEM, Cassart J, Coche T, Vinals Y De Baesolsc
TX

```

XX WPI; 2000-572268/53.
DR P-PSDB; AAB26325.
DR

PT New human CASB618 polypeptide, useful as a vaccine for prophylactic and
PT therapeutic treatment of cancers, particularly ovarian or colon cancer,
PT autoimmune diseases and related conditions.

PS Claim 13; Page 61; 76pp; English.

CC The present sequence is the coding sequence of human CASB618 protein. The
CC gene for human CASB618 is thought to be located on chromosome 15. The
CC protein encoded by the present sequence and epitopes of the CASB618
CC protein (see AB231327 to AB261999) are useful in diagnosing the occurrence
CC of tumour cells and in vaccines for prophylactic and therapeutic
CC treatment of cancers, particularly ovarian or colon cancer, autoimmune
CC diseases and related conditions

QX Sequence 1441 BP; 285 A; 466 C; 392 G; 294 T; 0 U; 0 Other;

Query Match	Score	DB 3	Length
85.3%	425		1441

Best Local Similarity 100.0%; Pred. No. 1,1e-201;
Matches 425; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY	62	GGCGCGCCTTTCGAGGTACACGTGGCAACGGGCGCTCGTGTGCTCTTTCCTCGAAGGGAGGCG	12
Db	998	GGCGCGCCTTTCGAGGTACACGTGGCAACGGGCGCTCGTGTGCTCTTTCCTCGAAGGGAGGCG	105
QY	122	TGTGAGTCTCCAGATATGTTGGGCCAGCGCTCTTGCAACCTTTCGACCAAGCGCA	181
Db	1058	TGTGAGTCTCCAGATATGTTGGGCCAGCGCTCTTGCAACCTTTCGACCAAGCGCA	1117
QY	182	AGGACTGCAGCAGGAGAGAGGGGGCTCACTCTTATCCTCGGCAACCACTGCAACAGC	241
Db	1118	AGGACTGCAGCAGGAGAGAGGGGGCTCACTCTTATCCTCGGCAACCACTGCAACAGC	1177
QY	242	AGGCGCGCTCCAGACTTAAATATGATACACACTAACTGTAGGGGGAGCCCAATCTGG	301
Db	1178	AGGCGCGCTCCAGACTTAAATATGATACACACTAACTGTAGGGGGAGCCCAATCTGG	123

Qy	302	ACTCCTTCCCGCCCTTGGGACATCGAGAGCCGGGAAAGAGTGCACCGCCAGAGGCTGGAGCA	361
Db	1238	ACTCCTTCCCGCCCTTGGGACATCGAGAGCCGGGAAAGAGTGCACCGCCAGAGGCTGGAGCA	1297
Qy	362	GGAGAGCTTCAGGAAGGCACTGAGCGCTGTGTGGCGAGGCTTCGGAATCGCAGGCA	421
Db	1298	GGAGAGCTTCAGGAAGGCACTGAGCGCTGTGTGGCGAGGCTTCGGAATCGCAGGCA	1357
Qy	422	CCAGGGAAGTCTCTTGGGGCGATCTGTAAATAAAGCTTTTTTCTTTTGTTTTTTAAAA	481
Db	1358	CCAGGGAAGTCTCTTGGGGCGATCTGTAAATAAAGCTTTTTTCTTTTGTTTTTTAAAA	1417
Qy	482	AAAAA 486	
Db	1418	AAAAA 1422	

RESULT 3
ABL90535
ID ABL90535 standard; cDNA; 1474 BP.
AC
XX ABL90535;
XX
DT 24-MAY-2002 (first entry)
XX
XX
DE Human polynucleotide SEQ ID NO 1097.
XX
XX Cytostatic; immunosuppressive; nocotropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; gene; ss.
XX

OS Homo sapiens.
 XX W0200190304-A2.
 XX 29-NOV-2001.
 XX 18-MAY-2001; 2001MO-US016450.
 XX 19-MAY-2000; 2000US-0205515P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Birse CE, Rosen CA;
 XX WPI: 2002-122018/16.
 XX P-PSDB; ABB90126.
 XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
 PT prevention of neural, immune system, muscular, reproductive,
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
 PT disorders.
 XX Claim 4; SEQ ID NO 1097; 2081bp + Sequence listing; English.
 XX The invention relates to novel genes (ABE9449-ABU9083) and proteins
 CC (ABB9040-ABB9044) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
 CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
 CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
 CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
 CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
 CC infectious diseases such as viral, bacterial, fungal and parasitic
 CC infections. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX Sequence 1474 BP; 292 A; 480 C; 404 G; 296 T; 0 U; 2 Other;
 SQ
 Query Match 85.3%; Score 425; DB 6; Length 1474;
 Best Local Similarity 100.0%; Pred. No. 1.1e-201;
 Matches 425; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 62 GCACGCGCTTCTGAGTCAAGTGGCAACGCGCGTCTGAGTCTCTCTGAGGAGGCGG 121
 Db 1024 GCGCGCGCTTCTGAGTCAAGTGGCAACGCGCGTCTGAGTCTCTCTGAGGAGGCGG 1083
 Oy 122 TGGTGAAGTCTTCAATATGTTGGCCAGCCCTCTTGGACCTTCTGACCAAGCGCA 181
 Db 1084 TGGTGAAGTCTTCAATATGTTGGCCAGCCCTCTTGGACCTTCTGACCAAGCGCA 1143
 Oy 182 AGGAGTGGAGCCAG 241
 Db 1144 AGGAGTGGAGCCAG 1203
 Oy 242 AGGCGCGCTTCTGAGTCAAGTGGCAACGCGCGTCTGAGTCTCTCTGAGGAGGCGG 301
 Db 1204 AGGCGCGCTTCTGAGTCAAGTGGCAACGCGCGTCTGAGTCTCTCTGAGGAGGCGG 1263
 Oy 302 ACTGCTTCCCGGCTTGGAGCATGCGAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 361
 Db 1264 ACTGCTTCCCGGCTTGGAGCATGCGAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAG 1323
 Oy 362 GGAAGAGTCCAG 421
 Db 1324 GGAAGAGTCCAG 1383
 Oy 422 CCAAGGAAAGTCTCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 481

Db 1384 CCAAGGAAAGTCTCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1443
 Oy 482 AAAA 486
 Db 1444 AAAA 1448
 RESULT 4
 ADD19228
 ID ADD19228 standard; cDNA; 1491 BP.
 XX AC ADD19228;
 XX 15-JAN-2004 (first entry)
 XX Human cDNA from secreted protein gene 45.
 XX human secreted protein; cytosolic; antibacterial; virucide;
 KW neuroprotective; gynaecological; gastrointestinal-gen; cardiant;
 KW cardiovascular-gen; nephrotropic; antiinflammatory; muscular-gen;
 KW respiratory-gen; immunosuppressive; cerebroprotective; vasotropic;
 KW neotropic; anti-allergic; cancer; bacterial infection; viral infection;
 KW neural disorder; immune system disorder; blood disorder;
 KW muscular disorder; reproductive disorder; gastrointestinal disorder;
 KW pulmonary disorder; cardiovascular disorder; renal disorder;
 KW inflammatory disorder; proliferative disorder; human; ss; gene.
 OS Homo sapiens.
 XX W02003052377-A2.
 XX 26-JUN-2003.
 XX 06-NOV-2002; 2002MO-US035606.
 XX 07-NOV-2001; 2001US-0331046P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Ruben SM;
 XX WPI: 2003-533050/50.
 XX P-PSDB; ADD19303.
 XX New isolated nucleic acids encoding signal transduction pathway component
 PT polypeptides, useful for diagnosing, treating, and/or preventing
 PT disorders, such as cancer, infections, cardiovascular and inflammatory
 PT diseases.
 XX Claim 1; SEQ ID NO 55; 554bp; English.
 XX The invention relates to an isolated nucleic acid molecule (cDNA)
 CC encoding a human secreted protein, representing one of 85 novel genes.
 CC Also included are recombinant vectors, host cells (expressing the
 CC protein), the secreted proteins (including their fragments, epitopes and
 CC homologues), an isolated antibody that binds specifically to the protein,
 CC a homologue of a pathological condition or susceptibility to a pathological
 CC condition (comprising determining the presence or absence of a mutation
 CC in the nucleic acid and diagnosing a pathological condition or
 CC absence of the mutation), diagnosing a pathological condition or
 CC susceptibility to a pathological condition (comprising determining the
 CC presence or amount of expression of the protein in a biological sample
 CC and diagnosing a condition based on the presence or amount of expression
 CC of the protein), preventing, treating or ameliorating a medical condition
 CC by administering the nucleic acid or protein to a mammalian subject,
 CC identifying a binding partner to the protein, the gene corresponding to
 CC the cDNA sequence, and identifying an activity in a biological assay
 CC (comprising expressing the nucleic acid in a cell, isolating the
 CC supernatant, detecting an activity in a biological assay and identifying
 CC the protein in the supernatant having the activity). The nucleic acids
 CC and proteins display the following activities: Cytostatic, antibacterial,
 CC virucide, Neuroprotective, Gynaecological, Gastrointestinal-gen,

CC Cardiant, Cardiovascular-Gen, Nephrotoxic, Antiinflammatory, Muscular-
 CC Gen, Respiratory-Gen, Immunosuppressive, Cerebroprotective, Vasotropic,
 CC Neurotic, Antiallergic. The methods and compositions of the present
 CC invention are useful for diagnosing, treating, preventing and/or
 CC prognosticating disorders related to the novel polypeptides, such as
 CC cancer, bacterial or viral infections, and neural, immune system, blood,
 CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
 CC renal, inflammatory or proliferative disorders (many examples of these
 CC diseases and disorders are given in the specification). The present
 CC sequence encodes a novel secreted protein of the invention.

XX
 XX Sequence 1491 BP; 293 A; 485 C; 410 G; 302 T; 0 U; 1 Other;
 SQ

Query Match 85.3%; Score 425; DB 9; Length 1491;
 Best Local Similarity 100.0%; Pred. No. 1.1e-201;
 Matches 425; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 GCGCGCCCTTTGGGTCACGCTGGCAACCGGCTCTTCTCTTCTCTGAGGCGCG 121
 DB 1041 GCGCGCCCTTTGGGTCACGCTGGCAACCGGCTCTTCTCTTCTCTGAGGCGCG 1100

QY 122 TGGTAGCTCCAGATGTTTCGCGCCAGCGCTCTTCCGACCTTCTGACCAAGCGCA 181
 DB 1101 TGGTAGCTCCAGATGTTTCGCGCCAGCGCTCTTCCGACCTTCTGACCAAGCGCA 1160

QY 182 AGGACTGCAGCCAG 241
 DB 1161 AGGACTGCAGCCAG 1220

QY 242 AGGCGCGCTCTCCAGACTTAAATGATACCACTACCTGAGAGAGAGAGAGAGAGAG 301
 DB 1221 AGGCGCGCTCTCCAGACTTAAATGATACCACTACCTGAGAGAGAGAGAGAGAGAG 1280

QY 302 ACTCTCTTCCCGCTTGGAGCATTCGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAG 361
 DB 1281 ACTCTCTTCCCGCTTGGAGCATTCGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAG 1340

QY 362 GAGAGAGCTCCAG 421
 DB 1341 GAGAGAGCTCCAG 1400

QY 422 CCAGGAAAGTCTCTCTGAGGCGATCTGTAAATAAACCTTTTCTTTTGTATTTTAA 481
 DB 1401 CCAGGAAAGTCTCTCTGAGGCGATCTGTAAATAAACCTTTTCTTTTGTATTTTAA 1460

QY 482 AAAA 486
 DB 1461 AAAA 1465

RESULT 5
 ID AAA96505 standard; cDNA; 1421 BP.
 XX
 XX AAA96505;
 AC
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE cDNA encoding a human transmembrane protein.
 XX
 KW Human, transmembrane protein; cell proliferation disorder; myeloma;
 KW reproductive disorder; smooth muscle disorder; neurological disorder;
 KW arteriosclerosis; leukaemia; acquired immunodeficiency syndrome; AIDS;
 KW allergy; ovulatory defect; angina; hypertension; stroke; epilepsy;
 KW Alzheimer's disease; Tourette's disorder; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT 267..1229
 FT CDS /*tag= a
 XX
 FN WO20056891-A2.

XX
 XX 28-SEP-2000.
 XX
 PF 22-MAR-2000; 2000WO-US007817.
 XX
 XX 22-MAR-1999; 99US-0125537P.
 PR 16-JUN-1999; 99US-0139565P.
 XX
 XX (INCY-) INCYTE PHARM INC.
 PA
 XX Yue H, Lal P, Tang YT, Hillman JL, Reddy R, Bandman O,
 PI Baughn WR, Lu DM, Azimzai Y, Yang J;
 DR WPI; 2000-579485/54.
 XX P-PSDB; AAB18992.
 DR

PT New human transmembrane proteins are used to treat a disease or condition
 PT associated with decreased expression of functional HTMP e.g. Tourette's
 PT disorder, angina and leukemia.
 PS
 PS Claim 4; Page 129; 130pp; English.

CC The present sequence encodes a human transmembrane proteins (HTMP).
 CC Agonists and antagonists of the protein are used to treat a disease or
 CC conditions associated with overexpression of the protein. Diseases and
 CC conditions which can be treated include cell proliferative,
 CC immunological, reproductive, smooth muscle and neurological disorders
 CC e.g. arteriosclerosis, myeloma, leukaemia, acquired immunodeficiency
 CC syndrome (AIDS), allergies, ovulatory defects, angina, hypertension,
 CC stroke, Alzheimer's disease, epilepsy and Tourette's disorder. The
 CC polynucleotides may be used to detect and quantify gene expression in
 CC biopsied tissues where protein expression may be correlated with disease
 CC e.g. to determine absence, presence or excess expression of HTMP or to
 CC monitor regulation of HTMP expression during therapeutic intervention
 CC
 XX

SQ Sequence 1421 BP; 264 A; 468 C; 395 G; 294 T; 0 U; 0 Other;

Query Match 83.5%; Score 416; DB 3; Length 1421;
 Best Local Similarity 100.0%; Pred. No. 3.4e-197;
 Matches 416; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 GCGCGCCCTTTGGGTCACGCTGGCAACCGGCTCTTCTCTTCTCTGAGGCGCG 121
 DB 1006 GCGCGCCCTTTGGGTCACGCTGGCAACCGGCTCTTCTCTTCTCTGAGGCGCG 1065

QY 122 TGGTAGCTCCAGATGTTTCGCGCCAGCGCTCTTCCGACCTTCTGACCAAGCGCA 181
 DB 1066 TGGTAGCTCCAGATGTTTCGCGCCAGCGCTCTTCCGACCTTCTGACCAAGCGCA 1125

QY 182 AGGACTGCAGCCAG 241
 DB 1126 AGGACTGCAGCCAG 1185

QY 242 AGGCGCGCTCTCCAGACTTAAATGATACCACTACCTGAGAGAGAGAGAGAGAGAG 301
 DB 1186 AGGCGCGCTCTCCAGACTTAAATGATACCACTACCTGAGAGAGAGAGAGAGAGAG 1245

QY 302 ACTCTCTTCCCGCTTGGAGCATTCGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAG 361
 DB 1246 ACTCTCTTCCCGCTTGGAGCATTCGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAG 1305

QY 362 GAGAGAGCTCCAG 421
 DB 1306 GAGAGAGCTCCAG 1365

QY 422 CCAGGAAAGTCTCTCTGAGGCGATCTGTAAATAAACCTTTTCTTTTGTATTTT 477
 DB 1366 CCAGGAAAGTCTCTCTGAGGCGATCTGTAAATAAACCTTTTCTTTTGTATTTT 1421

RESULT 6
 ID ADA53627/c
 ID ADA53627 standard; cDNA; 2684 BP.


```

Db      3380 GGGGTCCTGCTCTTCTTCCGAGGGCCGCGTGAATCTCCAGTATGTTCCGCCAC 3439
QY      151 GCTCTTGGACCCCTTCTGAGCAAAAGCCGCAAGACTGACCCAGAGAGAGGGGCTCA 210
Db      3440 GCTCTTGGACCCCTTCTGAGCAAAAGCCGCAAGACTGACCCAGAGAGAGGGGCTCA 3499
QY      211 CCTTATCTCTGGGAGACCACTGACAAAGAGGGCGCTCCCAAGCTTAAATGATC 270
Db      3500 CCTTATCTCTGGGAGACCACTGACAAAGAGGGCGCTCCCAAGCTTAAATGATC 3559
QY      271 ACCACTAACCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCTTGGACATCGAGG 330
Db      3560 ACCACTAACCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCTTGGACATCGAGG 3619
QY      331 CCGGGAAGAGAGTCCCGGAGGCTGAGGCGAGAGAGTCCAGAGAGAGGCACTGAGGCT 390
Db      3620 CCGGGAAGAGAGTCCCGGAGGCTGAGGCGAGAGAGTCCAGAGAGAGGCACTGAGGCT 3679
QY      391 GCTGCGCGAGAGGCTTGGACATCCGAGGCAACAGGAAAGTCTCTGGGCGATCTGTA 450
Db      3680 GCTGCGCGAGAGGCTTGGACATCCGAGGCAACAGGAAAGTCTCTGGGCGATCTGTA 3739
QY      451 AATTAACCTTTTCTTTTCTTTTAAAAA 482
Db      3740 AATTAACCTTTTCTTTTCTTTTAAAAA 3771

RESULT 8
ID      ABZ74587 standard; DNA; 5033 BP.
AC      ABZ74587;
DT      12-MAY-2003 (first entry)
DE      Secreted protein gene 360 genomic fragment HUFCL31, SEQ ID NO:1734.
XX      XX
XX      Human; secreted protein; cancer; tumour; hyperproliferative disorder;
XX      autoimmune disorder; inflammation; angiogenic diseases; AIDS;
XX      acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
XX      drug screening; chromosome identification; chromosome mapping;
XX      cytotoxic; gene therapy; anti-inflammatory; immunomodulator; anti-HIV;
XX      antineoplastic; vulnereary; chromosome 9p21; gene; ds.
OS      Homo sapiens.
XX      XX
XX      WO200277013-A2.
XX      XX
XX      03-OCT-2002.
XX      XX
XX      26-MAR-2002; 2002MO-US009370.
XX      XX
XX      27-MAR-2001; 2001US-0278650P.
XX      PR      12-SEP-2001; 2001US-00950082.
XX      PR      12-SEP-2001; 2001US-00950083.
XX      XX
XX      (HUMA-) HUMAN GENOME SCI INC.
XX      XX
XX      PI      Rosen CA, Ruben SM;
XX      DR      WPI; 2003-040578/03.
XX      XX
XX      New human secreted proteins and nucleic acids, useful for detecting or
XX      PT      treating cancer or other hyperproliferative disorders, autoimmune
XX      PT      disorders, inflammatory disorders, HIV disease, hepatitis or anaemia.
XX      PS      Disclosure; Page 2315-2316; 2474pp; English.
XX      XX
XX      ABZ73381-ABZ73697 represent cDNAs corresponding to 391 human secreted
XX      CC      protein genes, and ABZ0947-ABP01363 represent the proteins they encode.
XX      CC      ABZ73698-ABZ74687 represent human secreted protein genomic fragments. The
XX      CC      invention also encompasses antibodies specific for the secreted proteins,

```

```

CC      the use of the secreted proteins in drug screening and recombinant
CC      CC      vectors and host cells comprising a nucleic acid of the invention. The
CC      CC      secreted proteins are thought to be involved in biological activities
CC      CC      associated with cellular signalling, cellular differentiation, cell
CC      CC      migration, prothrombin activation and neurotransmitter activity. The
CC      CC      secreted proteins, nucleic acids encoding them, antibodies or antibody
CC      CC      fragments specific for the secreted proteins, and modulators of protein
CC      CC      activity are useful for the secreted proteins, and modulators of protein
CC      CC      hyperproliferative disorders. Additionally, the secreted proteins and
CC      CC      their nucleic acids may also be used in the treatment of autoimmune
CC      CC      disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
CC      CC      (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
CC      CC      wound healing. Nucleic acids of the invention may be used for chromosome
CC      CC      identification, chromosome mapping, in gene therapy, for identifying
CC      CC      individuals from minute biological samples, as hybridisation probes, and
CC      CC      as molecular weight markers. The present sequence represents a human
CC      CC      secreted protein genomic fragment referred to in the disclosure of the
CC      CC      invention
XX      XX
SQ      Sequence 5033 BP; 1061 A; 1409 C; 1465 G; 1098 T; 0 U; 0 Other;
Query Match      78.7%; Score 392; DB 7; Length 5033;
Best Local Similarity 100.0%; Pred. No. 3.4e-185;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      91 GGGTCCTGCTGCTCTTCTTCTGAGAGGGCCGCTGAGTCTCCAGTATGTTCCGCCAC 150
Db      3380 GGGTCCTGCTGCTCTTCTTCTGAGAGGGCCGCTGAGTCTCCAGTATGTTCCGCCAC 3439
QY      151 GCTCTTGGACCCCTTCTGAGCAAAAGCCGCAAGACTGACCCAGAGAGAGGGGCTCA 210
Db      3440 GCTCTTGGACCCCTTCTGAGCAAAAGCCGCAAGACTGACCCAGAGAGAGGGGCTCA 3499
QY      211 CCTTATCTCTGGGAGACCACTGACAAAGAGGGCGCTCCCAAGCTTAAATGATC 270
Db      3500 CCTTATCTCTGGGAGACCACTGACAAAGAGGGCGCTCCCAAGCTTAAATGATC 3559
QY      271 ACCACTAACCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCTTGGACATCGAGG 330
Db      3560 ACCACTAACCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCTTGGACATCGAGG 3619
QY      331 CCGGGAAGAGAGTCCCGGAGGCTGAGGCGAGAGAGTCCAGAGAGAGGCACTGAGGCT 390
Db      3620 CCGGGAAGAGAGTCCCGGAGGCTGAGGCGAGAGAGTCCAGAGAGAGGCACTGAGGCT 3679
QY      391 GCTGCGCGAGAGGCTTGGACATCCGAGGCAACAGGAAAGTCTCTGGGCGATCTGTA 450
Db      3680 GCTGCGCGAGAGGCTTGGACATCCGAGGCAACAGGAAAGTCTCTGGGCGATCTGTA 3739
QY      451 AATTAACCTTTTCTTTTCTTTTAAAAA 482
Db      3740 AATTAACCTTTTCTTTTCTTTTAAAAA 3771

RESULT 9
ADCC1005
ID      ADCC1005 standard; DNA; 5033 BP.
AC      ADCC1005;
DT      18-DEC-2003 (first entry)
DE      Human secreted protein-related DNA sequence #423.
XX      XX
XX      Gene therapy; human; secreted protein; haemopoietic disorder;
XX      XX      haematological disorder; anaemia; haemophilia; inflammatory disorder;
XX      XX      inflammatory bowel disease; Crohn's disease; neoplastic disease; cancer;
XX      XX      leukemia; wound healing; epithelial cell proliferation disorder;
XX      XX      immune disorder; autoimmune disorder; asthmatic disorder;
XX      XX      cardiovascular disorder; atherosclerosis; myocarditis;
XX      XX      infectious disease; HIV; AIDS; endocrine disorder; diabetes;
XX      XX      gastrointestinal disorder; duodenal ulcer; gastroenteritis; gene; ds.

```

OS Homo sapiens.
 XX MO200292787-A2.
 XX 21-NOV-2002.
 XX 26-MAR-2002; 2002WO-US009257.
 XX 27-MAR-2001; 2001US-0278650P.
 XX 12-SEP-2001; 2001US-00950082.
 XX 12-SEP-2001; 2001US-00950083.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Ruben SM;
 XX WPI; 2003-129287/12.
 XX
 PT New human secreted proteins and nucleic acid molecules, useful for
 PT preparing a diagnostic or pharmaceutical composition for diagnosing,
 PT preventing or treating hematopoietic or hematologic disorders, e.g.
 PT anemia or hemophilia.
 XX
 XX Disclosure; SEQ ID NO 959; 1512pp; English.
 XX
 CC The invention comprises the amino acid and coding sequences of human
 CC secreted proteins. The DNA and protein sequences of the invention are
 CC useful for detecting, preventing, diagnosing, prognosticating, treating
 CC or ameliorating: haematopoietic or haematological disorders (e.g. anaemia
 CC and hemophilia); inflammatory disorders (e.g. inflammatory bowel disease
 CC and Crohn's disease); neoplastic disease (e.g. cancer and leukaemia);
 CC wound healing and disorders of epithelial cell proliferation; immune
 CC disorders (e.g. autoimmune disorders and asthmatic disorders);
 CC cardiovascular disorders (e.g. atherosclerosis and myocarditis);
 CC infectious disease (e.g. HIV/AIDS); endocrine disorders (e.g. diabetes);
 CC and gastrointestinal disorders (e.g. duodenal ulcers and
 CC gastroenteritis). The present DNA sequence was used in the
 CC exemplification of the invention.
 CC
 XX Sequence 5033 BP; 1061 A; 1409 C; 1465 G; 1098 T; 0 U; 0 Other;
 SO

Query Match 78.7%; Score 392; DB 9; Length 5033;
 Best Local Similarity 100.0%; Pred. No. 3.4e-185;
 Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 GAGCGTCTGCTGCTCTTCTTCTGAGAGGCGGTGTGATCTCCAGTATGTTGGCCGAGC 150
 DB 3380 GAGTCTCTGCTCTTCTTCTGAGAGGCGGTGTGATCTCCAGTATGTTGGCCGAGC 3439
 QY 151 GCTCTTGACACCTTTCTTGACCAAGCGCCCAAGAGCTGACCCAGAGAGAGGGGGCTCA 210
 DB 3440 GCTCTTGACACCTTTCTTGACCAAGCGCCCAAGAGCTGACCCAGAGAGAGGGGGCTCA 3499
 QY 211 CCTTTATCTCGGGGACCACTGCAACAAGAGCGGCTCTCCAGCTTAAATATATC 270
 DB 3500 CCTTTATCTCGGGGACCACTGCAACAAGAGCGGCTCTCCAGCTTAAATATATC 3559
 QY 271 ACCACTTAACCTGTAGAGGGGAGCCCAATCTGACTCTTCCCGCTTGGAGCATCCAGG 330
 DB 3560 ACCACTTAACCTGTAGAGGGGAGCCCAATCTGACTCTTCCCGCTTGGAGCATCCAGG 3619
 QY 331 CCGGGAAGCAGTGGCCCGGAGAGGCTGGGCGGAGAGCTCCAGAAAGGCACTGACCGCT 390
 DB 3620 CCGGGAAGCAGTGGCCCGGAGAGGCTGGGCGGAGAGCTCCAGAAAGGCACTGACCGCT 3679
 QY 391 GCTGCGGAGAGGCTCGAGCATCCGACAGGACCAAGGAAAGTCTCTGGGGCATCTGTA 450
 DB 3680 GCTGCGGAGAGGCTCGAGCATCCGACAGGACCAAGGAAAGTCTCTGGGGCATCTGTA 3739
 QY 451 AATAAACCCTTTTCTTTTCTTTTAAAAA 482
 DB 3740 AATAAACCCTTTTCTTTTCTTTTAAAAA 3771

RESULT 10
 AAC95564
 ID AAC95564 standard; cDNA; 1460 BP.
 XX
 AC AAC95564;
 XX
 DT 21-FEB-2001 (first entry)
 XX
 DE Human secreted protein gene 44 SEQ ID NO:54.
 XX
 KM Human; secreted protein; cytosolic; immunosuppressive; neutrotropic;
 KM neutrotropic; antiviral; antiallergic; hepatotropic; antidiabetic;
 KM antiinflammatory; antiviral; anticonvulsant; antibacterial;
 KM antifungal; antiparasitic; cardiant; cancer; immune disease; allergy;
 KM cardiovascular disorder; wound healing; infection; neurological disease;
 KM ss.
 XX
 OS Homo sapiens.
 XX
 XX WO200061596-A1.
 XX
 XX 19-OCT-2000.
 XX
 XX 06-APR-2000; 2000WO-US008983.
 XX
 XX 09-APR-1999; 99US-0128703P.
 XX 14-JAN-2000; 2000US-0176068P.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX (ROSE/) ROSEN C A.
 XX
 XX Ruben SM, Komatsoulis G;
 XX WPI; 2000-611865/58.
 XX P-PSDB; AAB52055.
 XX
 PT Fifty nucleic acid molecules encoding human secreted proteins, useful in
 PT the prevention, treatment and diagnosis of cancer, immune disorders,
 PT cardiovascular disorders and neurological diseases.
 PT
 XX
 XX Claim 1; Page 443; 505pp; English.
 PS
 XX Polynucleotide sequences AAC95521 - AAC95570 represent cDNA encoding 50
 XX human secreted proteins AAB52012 - AAB52161. Sequences AAB52062 -
 XX AAC952103 represent alternative polypeptides encoded by the genes, and
 XX amino acid sequences with which they share homology. The genes and
 XX proteins have activities dependent on the tissues and cells in which they
 XX are expressed. Examples of their activities include cytostatic;
 XX immunosuppressive; neutrotropic; neuroprotective; antiviral; antiallergic;
 XX hepatotropic; antidiabetic; antiinflammatory; antiviral; vulnary;
 XX anticonvulsant; antibacterial; antifungal; antiparasitic; and cardiant.
 XX The secreted proteins, polypeptides, antenons and agonists may be
 XX useful in treating, preventing and/or diagnosing diseases and disorders
 XX such as cancer, particularly breast and ovarian cancer, and other cancers
 XX of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract,
 XX liver, lung, or urogenital. Immune disorders such as Addison's disease,
 XX allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 XX diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 XX arthritis and ulcerative colitis; cardiovascular disorders such as
 XX myocardial ischaemia; wound healing; neurological diseases such as
 XX cerebral anoxia and epilepsy; and infectious diseases such as viral,
 XX bacterial, fungal and parasitic infections may also be treated using the
 XX proteins and polynucleotides of the invention. Sequences AAC95512 -
 XX AAC95520 and AAB52011 are used in the isolation and characterisation of
 XX the proteins and polynucleotides of the invention
 SO

Query Match 45.8%; Score 228; DB 3; Length 1460;
 Best Local Similarity 100.0%; Pred. No. 2.2e-103;
 Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 259 TTTAAATGTATCACCACCTAAGCTGTAGAGGGGAGCCCAATCTGACTCTTCCCGCCTTG 318
 DB 1217 TTTAAATGTATCACCACCTAAGCTGTAGAGGGGAGCCCAATCTGACTCTTCCCGCCTTG 1276
 QY 319 GGACATCGCAGCGCGGAGAGCAGTGCCTCCGCAAGCCTTGGCCAGAGAGCTCCAGAGAG 378
 DB 1277 GGACATCGCAGCGCGGAGAGCAGTGCCTCCGCAAGCCTTGGCCAGAGAGCTCCAGAGAG 1336
 QY 379 GCACCTGAGCGCTGTGCGCGCAGGCTCGACATCCGAGGACCAAGAGAAAGTCTCTG 438
 DB 1337 GCACCTGAGCGCTGTGCGCGCAGGCTCGACATCCGAGGACCAAGAGAAAGTCTCTG 1396
 QY 439 GGGCGATCTGTAAATTAACCTTTTCTTTTCTTTTAAAAA 486
 DB 1397 GGGCGATCTGTAAATTAACCTTTTCTTTTCTTTTAAAAA 1444

RESULT 11

ABZ67235
 ID ABZ67235 standard, cDNA, 1460 BP.
 XX AC ABZ67235;
 XX 26-MAR-2003 (first entry)
 DE Human secreted protein encoding cDNA SEQ ID NO 355.

XX Human; secreted protein; nootropic; neuroprotective; cytostatic;
 KW vintidex; dermatological; immunosuppressive; anti-infective; anti-HIV;
 KW vintidex; dermatological; immunosuppressive; anti-infective; anti-HIV;
 KW anti-infective; cancer; antineoplastic; hepatocellular carcinoma;
 KW anti-infective; cancer; antineoplastic; hepatocellular carcinoma;
 KW antifungal; antiparasitic; cardiac; immune disorder; infection; vaccine;
 KW cardiovascular disorder; neurological disease; neurotropic;
 KW gene therapy; gene; chromosome 9p21; ds.

XX Homo sapiens.
 PN WO200277186-A2.
 XX 03-OCT-2002.
 PD 26-MAR-2002; 2002WO-US009188.
 PF 27-MAR-2001; 2001US-0278650P.
 PR 12-SEP-2001; 2001US-00950082.
 PR 12-SEP-2001; 2001US-00950083.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Ruben SM;
 PI WPI; 2003-040583/03.
 DR P-PSDB; ABP99814.
 PT New human secreted proteins encoded by genes contained in cDNA clones
 PT (e.g. HGCAC19), useful for preventing, treating or diagnosing e.g. AIDS,
 PT multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or
 PT West Nile fever.

PS Claim 7; Page 1353; 2423pp; English.

XX The invention relates to novel human genes (ABZ6691-ABZ68209) and the
 CC encoded secreted proteins (ABP99470-ABP99872) useful for preventing,
 CC treating or ameliorating medical conditions e.g. by protein or gene
 CC therapy. The genes are isolated from a range of human tissues disclosed
 CC in the specification. The nucleic acids, proteins, antibodies and
 CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
 CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the
 CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
 CC lung or urogenital; (b) immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid

CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
 CC myocardial ischemias; (d) wound healing; (e) neurological diseases e.g.
 CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
 CC bacterial, fungal and parasitic infections
 XX Sequence 1460 BP; 292 A; 472 C; 398 G; 298 T; 0 U; 0 Other;

Query Match 45.8%; Score 228; DB 7; Length 1460;
 Best Local Similarity 100.0%; Pred. No. 2,2e-103;
 Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 259 TTTAAATGTATCACCACCTAAGCTGTAGAGGGGAGCCCAATCTGACTCTTCCCGCCTTG 318
 DB 1217 TTTAAATGTATCACCACCTAAGCTGTAGAGGGGAGCCCAATCTGACTCTTCCCGCCTTG 1276
 QY 319 GGACATCGCAGCGCGGAGAGCAGTGCCTCCGCAAGCCTTGGCCAGAGAGCTCCAGAGAG 378
 DB 1277 GGACATCGCAGCGCGGAGAGCAGTGCCTCCGCAAGCCTTGGCCAGAGAGCTCCAGAGAG 1336
 QY 379 GCACCTGAGCGCTGTGCGCGCAGGCTCGACATCCGAGGACCAAGAGAAAGTCTCTG 438
 DB 1337 GCACCTGAGCGCTGTGCGCGCAGGCTCGACATCCGAGGACCAAGAGAAAGTCTCTG 1396
 QY 439 GGGCGATCTGTAAATTAACCTTTTCTTTTCTTTTAAAAA 486
 DB 1397 GGGCGATCTGTAAATTAACCTTTTCTTTTCTTTTAAAAA 1444

RESULT 12

ABZ73640
 ID ABZ73640 standard, cDNA, 1460 BP.

XX ABZ73640;
 DE 12-MAY-2003 (first entry)

XX Secreted protein-encoding gene 360 cDNA clone HUC131, SEQ ID NO:370.
 DE 12-MAY-2003 (first entry)
 XX Human; secreted protein; cancer; tumour; hyperproliferative disorder;
 KW autoimmune disorder; inflammation; angiotensin; anaemia; wound healing;
 KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
 KW drug screening; chromosome identification; chromosome mapping;
 KW cytostatic; gene therapy; anti-infective; immunomodulator; anti-HIV;
 KW antineoplastic; vulnery; chromosome 9p21; gene; ss.

XX Homo sapiens.
 PN WO200277013-A2.
 XX 03-OCT-2002.
 PD 26-MAR-2002; 2002WO-US009370.
 PF 27-MAR-2001; 2001US-0278650P.
 PR 12-SEP-2001; 2001US-00950082.
 PR 12-SEP-2001; 2001US-00950083.
 XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;
 PI WPI; 2003-040578/03.
 DR P-PSDB; ABR01306.

XX New human secreted proteins and nucleic acids, useful for detecting or
 PT treating cancer or other hyperproliferative disorders, autoimmune
 PT disorders, inflammatory disorders, HIV disease, hepatitis or anaemia.
 XX Claim 21; Page 1345; 2474pp; English.

XX ABZ73281-ABZ73697 represent cDNAs corresponding to 391 human secreted
 CC protein genes, and ABP00947-ABP01363 represent the proteins they encode.
 CC ABZ73698-ABZ74687 represent human secreted protein genomic fragments. The

CC invention also encompasses antibodies specific for the secreted proteins,
CC the use of the secreted proteins in drug screening and recombinant
CC vectors and host cells comprising a nucleic acid of the invention. The
CC secreted proteins are thought to be involved in biological activities
CC associated with cellular signalling, cellular differentiation, cell
CC migration, prohormone activation and neurotransmitter activity. The
CC secreted proteins, nucleic acids encoding them, antibodies or antibody
CC fragments specific for the secreted proteins, and modulators of protein
CC activity are useful for diagnosing or treating cancers or other
CC hyperproliferative disorders. Additionally, the secreted proteins and
CC their nucleic acids may also be used in the treatment of autoimmune
CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
CC wound healing. Nucleic acids of the invention may be used for chromosome
CC identification, chromosome mapping, in gene therapy, for identifying
CC individuals from minute biological samples, as hybridisation probes, and
CC as molecular weight markers. The present sequence represents a human
CC secreted protein-encoding cDNA clone of the invention
XX
SQ Sequence 1460 BP; 292 A; 472 C; 398 G; 298 T; 0 U; 0 Other;
Query Match 45.8%; Score 228; DB 7; Length 1460;
Best Local Similarity 100.0%; Pred. No. 2.2e-103; Indels 0; Gaps 0;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 259 TTTAAATGTATCACCCTAATCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCTTG 318
DB 1217 TTTAAATGTATCACCCTAATCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCTTG 1276
QY 319 GGACATCGCAGCGCGGAGAGAGTCCCGCAGGCTCGGAGAGCTCCAGGAAG 378
DB 1277 GGACATCGCAGCGCGGAGAGAGTCCCGCAGGCTCGGAGAGCTCCAGGAAG 1336
QY 379 GCACATGAGCGCTGTGCGGAGGCTTCGACATCCGACGACCAAGGAAATCTCTCG 438
DB 1337 GCACATGAGCGCTGTGCGGAGGCTTCGACATCCGACGACCAAGGAAATCTCTCG 1396
QY 439 GGGGAGTCTGTAAATTAACCTTTTCTTTGTTTTTAAAAAAA 486
DB 1397 GGGGAGTCTGTAAATTAACCTTTTCTTTGTTTTTAAAAAAA 1444
RESULT 13
ID ADC20289 standard; DNA; 1460 BP.
AC ADC20289;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human secreted protein coding sequence #228.
XX
KW gene therapy; human; secreted protein; haemopoietic disorder;
KW haematological disorder; anaemia; haemophilia; inflammatory disorder;
KW inflammatory bowel disease; Crohn's disease; neoplastic disease; cancer;
KW leukaemia; wound healing; epithelial cell proliferation disorder;
KW immune disorder; autoimmune disorder; asthmatic disorder;
KW cardiovascular disorder; atherosclerosis; myocarditis;
KW infectious disease; HIV; AIDS; endocrine disorder; diabetes;
KW gastrointestinal disorder; duodenal ulcer; gastroenteritis; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200292787-A2.
XX
PD 21-NOV-2002.
XX
PF 26-MAR-2002; 2002WO-US009257.
XX
XX 27-MAR-2001; 2001US-0278650P.
PR 12-SEP-2001; 2001US-0095082.
PR 12-SEP-2001; 2001US-0095083.
XX

PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX WPI; 2003-129287/12.
XX
PT New human secreted proteins and nucleic acid molecules, useful for
PT preparing a diagnostic or pharmaceutical composition for diagnosing,
PT preventing or treating hematopoietic or hematologic disorders, e.g.
PT anemia or hemophilia.
XX
PS Claim 1; SEQ ID NO 238; 1513bp; English.
XX
XX The invention comprises the amino acid and coding sequences of human
XX secreted proteins. The DNA and protein sequences of the invention are
XX useful for detecting, preventing, diagnosing, prognosticating, treating
XX or ameliorating hematopoietic or haematological disorders (e.g. anaemia
XX and haemophilia); inflammatory disorders (e.g. inflammatory bowel disease
XX and Crohn's disease); neoplastic disease (e.g. cancer and leukaemia);
XX wound healing and disorders of epithelial cell proliferation; immune
XX disorders (e.g. autoimmune disorders and asthmatic disorders);
XX cardiovascular disorders (e.g. atherosclerosis and myocarditis);
XX infectious disease (e.g. HIV/AIDS); endocrine disorders (e.g. diabetes);
XX and gastrointestinal disorders (e.g. duodenal ulcers and
XX gastroenteritis). The present DNA sequence encodes a human secreted
XX protein of the invention.
SQ Sequence 1460 BP; 292 A; 472 C; 398 G; 298 T; 0 U; 0 Other;
Query Match 45.8%; Score 228; DB 9; Length 1460;
Best Local Similarity 100.0%; Pred. No. 2.2e-103; Indels 0; Gaps 0;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 259 TTTAAATGTATCACCCTAATCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCTTG 318
DB 1217 TTTAAATGTATCACCCTAATCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCTTG 1276
QY 319 GGACATCGCAGCGCGGAGAGAGTCCCGCAGGCTCGGAGAGCTCCAGGAAG 378
DB 1277 GGACATCGCAGCGCGGAGAGAGTCCCGCAGGCTCGGAGAGCTCCAGGAAG 1336
QY 379 GCACATGAGCGCTGTGCGGAGGCTTCGACATCCGACGACCAAGGAAATCTCTCG 438
DB 1337 GCACATGAGCGCTGTGCGGAGGCTTCGACATCCGACGACCAAGGAAATCTCTCG 1396
QY 439 GGGGAGTCTGTAAATTAACCTTTTCTTTGTTTTTAAAAAAA 486
DB 1397 GGGGAGTCTGTAAATTAACCTTTTCTTTGTTTTTAAAAAAA 1444
RESULT 14
ID AAK81950 standard; DNA; 525 BP.
AC AAK81950;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:36762.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001354.
PF 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR

PR	24-FEB-2000	2000US-01846560
PR	02-MAR-2000	2000US-01863504
PR	16-MAR-2000	2000US-01896744
PR	11-MAR-2000	2000US-01900764
PR	18-APR-2000	2000US-01981236
PR	19-MAY-2000	2000US-02055154
PR	07-JUN-2000	2000US-02056674
PR	28-JUN-2000	2000US-02148664
PR	30-JUN-2000	2000US-02151354
PR	07-JUL-2000	2000US-02166474
PR	07-JUL-2000	2000US-02168804
PR	11-JUL-2000	2000US-02174874
PR	11-JUL-2000	2000US-02174964
PR	14-JUL-2000	2000US-02183290
PR	26-JUL-2000	2000US-02200634
PR	26-JUL-2000	2000US-02200644
PR	14-AUG-2000	2000US-02245184
PR	14-AUG-2000	2000US-02245194
PR	14-AUG-2000	2000US-02252134
PR	14-AUG-2000	2000US-02252144
PR	14-AUG-2000	2000US-02252564
PR	14-AUG-2000	2000US-02255674
PR	14-AUG-2000	2000US-02256284
PR	14-AUG-2000	2000US-02255474
PR	14-AUG-2000	2000US-02254474
PR	14-AUG-2000	2000US-02257574
PR	14-AUG-2000	2000US-02257584
PR	14-AUG-2000	2000US-02257594
PR	18-AUG-2000	2000US-02262714
PR	18-AUG-2000	2000US-02266814
PR	22-AUG-2000	2000US-02266864
PR	22-AUG-2000	2000US-02271824
PR	30-AUG-2000	2000US-02270094
PR	30-AUG-2000	2000US-02298924
PR	01-SEP-2000	2000US-02292874
PR	01-SEP-2000	2000US-02293434
PR	01-SEP-2000	2000US-02293444
PR	01-SEP-2000	2000US-02293454
PR	05-SEP-2000	2000US-02295094
PR	05-SEP-2000	2000US-02295134
PR	06-SEP-2000	2000US-02304334
PR	06-SEP-2000	2000US-02304384
PR	08-SEP-2000	2000US-02311424
PR	08-SEP-2000	2000US-02311434
PR	08-SEP-2000	2000US-02311444
PR	08-SEP-2000	2000US-02311434
PR	08-SEP-2000	2000US-02311434
PR	08-SEP-2000	2000US-02320804
PR	12-SEP-2000	2000US-02320814
PR	14-SEP-2000	2000US-02331984
PR	14-SEP-2000	2000US-02332974
PR	14-SEP-2000	2000US-02332984
PR	14-SEP-2000	2000US-02332994
PR	14-SEP-2000	2000US-02332404
PR	14-SEP-2000	2000US-02332404
PR	14-SEP-2000	2000US-02330634
PR	14-SEP-2000	2000US-02330634
PR	14-SEP-2000	2000US-02330634
PR	21-SEP-2000	2000US-02334234
PR	21-SEP-2000	2000US-02342364
PR	21-SEP-2000	2000US-02342744
PR	25-SEP-2000	2000US-02349774
PR	25-SEP-2000	2000US-02349984
PR	26-SEP-2000	2000US-02354844
PR	27-SEP-2000	2000US-02356844
PR	27-SEP-2000	2000US-02356834
PR	29-SEP-2000	2000US-02356364
PR	29-SEP-2000	2000US-02356374
PR	29-SEP-2000	2000US-02356374
PR	29-SEP-2000	2000US-02356384
PR	29-SEP-2000	2000US-02356394
PR	29-SEP-2000	2000US-02356374
PR	02-OCT-2000	2000US-02370804
PR	02-OCT-2000	2000US-02370824
PR	02-OCT-2000	2000US-02370384

PR	02-OCT-2000;	2000US-0237639P.
PR	02-OCT-2000;	2000US-0237640P.
PR	13-OCT-2000;	2000US-0239935P.
PR	13-OCT-2000;	2000US-0239937P.
PR	20-OCT-2000;	2000US-0240960P.
PR	20-OCT-2000;	2000US-024121P.
PR	20-OCT-2000;	2000US-024185P.
PR	20-OCT-2000;	2000US-024186P.
PR	20-OCT-2000;	2000US-024187P.
PR	20-OCT-2000;	2000US-024180BP.
PR	20-OCT-2000;	2000US-024180PP.
PR	20-OCT-2000;	2000US-024182BP.
PR	01-NOV-2000;	2000US-0246117P.
PR	08-NOV-2000;	2000US-0246474P.
PR	08-NOV-2000;	2000US-0246475P.
PR	08-NOV-2000;	2000US-0246476P.
PR	08-NOV-2000;	2000US-0246477P.
PR	08-NOV-2000;	2000US-0246478P.
PR	08-NOV-2000;	2000US-0246523P.
PR	08-NOV-2000;	2000US-0246524P.
PR	08-NOV-2000;	2000US-0246525P.
PR	08-NOV-2000;	2000US-0246526P.
PR	08-NOV-2000;	2000US-0246527P.
PR	08-NOV-2000;	2000US-0246528P.
PR	08-NOV-2000;	2000US-0246532P.
PR	08-NOV-2000;	2000US-0246609P.
PR	08-NOV-2000;	2000US-0246610P.
PR	08-NOV-2000;	2000US-0246611P.
PR	08-NOV-2000;	2000US-0246613P.
PR	17-NOV-2000;	2000US-0249207P.
PR	17-NOV-2000;	2000US-0249208P.
PR	17-NOV-2000;	2000US-0249209P.
PR	17-NOV-2000;	2000US-0249210P.
PR	17-NOV-2000;	2000US-0249211P.
PR	17-NOV-2000;	2000US-0249212P.
PR	17-NOV-2000;	2000US-0249213P.
PR	17-NOV-2000;	2000US-0249214P.
PR	17-NOV-2000;	2000US-0249215P.
PR	17-NOV-2000;	2000US-0249216P.
PR	17-NOV-2000;	2000US-0249218P.
PR	17-NOV-2000;	2000US-0249244P.
PR	17-NOV-2000;	2000US-0249245P.
PR	17-NOV-2000;	2000US-0249264P.
PR	17-NOV-2000;	2000US-0249265P.
PR	17-NOV-2000;	2000US-0249297P.
PR	17-NOV-2000;	2000US-0249299P.
PR	17-NOV-2000;	2000US-0249300P.
PR	01-DEC-2000;	2000US-0250160P.
PR	01-DEC-2000;	2000US-0250391P.
PR	05-DEC-2000;	2000US-0251030P.
PR	05-DEC-2000;	2000US-0251988P.
PR	05-DEC-2000;	2000US-0256719P.
PR	06-DEC-2000;	2000US-0251479P.
PR	06-DEC-2000;	2000US-0251656P.
PR	08-DEC-2000;	2000US-0251688P.
PR	08-DEC-2000;	2000US-0251699P.
PR	08-DEC-2000;	2000US-0251899P.
PR	08-DEC-2000;	2000US-0251907P.
PR	11-DEC-2000;	2000US-0254097P.
PR	05-JAN-2001;	2001US-0259678P.
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
PI	Rosen CA, Barash SC, Ruben SM;	
XX		
DR	WPI, 2001-483426/52.	
XX		
PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides	
XX	useful for preventing, diagnosing and/or treating cancers and metastasis	
PS	Disclosure; SEQ ID NO 36762; 3071pp + Sequence Listing; English.	
XX		

CC AK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK67694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention

SQ Sequence 525 BP; 174 A; 76 C; 146 G; 129 T; 0 U; 0 Other;

Query Match

5.2%; Score 26; DB 4; Length 525;

Best Local Similarity 100.0%; Pred. No. 0.013; Mismatches 0; Gaps 0;

Matches 26; Conservative 0; Indels 0; Gaps 0;

QY 464 TTCTTTGTTTTTAAAAAATPA 489
DB 500 TTCTTTGTTTTTAAAAAATPA 525

RESULT 15

AAK81951
ID AAK81951 standard; DNA; 525 BP.

DT 07-NOV-2001 (first entry)

DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:36763.

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

KW cytostatic; gene therapy; vaccine; metastasis; ds.

OS Homo sapiens.

PN MO200157182-A2.

XX 09-AUG-2001.

PF 17-JAN-2001; 2001WO-US001354.

XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214986P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.

PR 14-AUG-2000; 2000US-0225759P.

PR 18-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.

PR 22-AUG-2000; 2000US-0226688P.

PR 22-AUG-2000; 2000US-0227182P.

PR 23-AUG-2000; 2000US-0227009P.

PR 30-AUG-2000; 2000US-0228924P.

PR 01-SEP-2000; 2000US-0229287P.

PR 01-SEP-2000; 2000US-0229343P.

PR 01-SEP-2000; 2000US-0229344P.

PR 01-SEP-2000; 2000US-0229345P.

PR 05-SEP-2000; 2000US-0229509P.

PR 05-SEP-2000; 2000US-0229513P.

PR 06-SEP-2000; 2000US-0230437P.

PR 08-SEP-2000; 2000US-0230438P.

PR 08-SEP-2000; 2000US-0231242P.

PR 08-SEP-2000; 2000US-0231243P.

PR 08-SEP-2000; 2000US-0231244P.

PR 08-SEP-2000; 2000US-0231413P.

PR 08-SEP-2000; 2000US-0231414P.

PR 08-SEP-2000; 2000US-0232080P.

PR 08-SEP-2000; 2000US-0232081P.

PR 12-SEP-2000; 2000US-0231968P.

PR 14-SEP-2000; 2000US-0232397P.

PR 14-SEP-2000; 2000US-0232398P.

PR 14-SEP-2000; 2000US-0232399P.

PR 14-SEP-2000; 2000US-0232400P.

PR 14-SEP-2000; 2000US-0232401P.

PR 14-SEP-2000; 2000US-0232402P.

PR 14-SEP-2000; 2000US-0233063P.

PR 14-SEP-2000; 2000US-0233064P.

PR 14-SEP-2000; 2000US-0233065P.

PR 21-SEP-2000; 2000US-0234223P.

PR 21-SEP-2000; 2000US-0234274P.

PR 25-SEP-2000; 2000US-0234997P.

PR 25-SEP-2000; 2000US-0234998P.

PR 26-SEP-2000; 2000US-0235484P.

PR 27-SEP-2000; 2000US-0235834P.

PR 27-SEP-2000; 2000US-0235836P.

PR 29-SEP-2000; 2000US-0236337P.

PR 29-SEP-2000; 2000US-0236338P.

PR 29-SEP-2000; 2000US-0236369P.

PR 29-SEP-2000; 2000US-0236370P.

PR 02-OCT-2000; 2000US-0236802P.

PR 02-OCT-2000; 2000US-0237037P.

PR 02-OCT-2000; 2000US-0237038P.

PR 02-OCT-2000; 2000US-0237039P.

PR 02-OCT-2000; 2000US-0237040P.

PR 13-OCT-2000; 2000US-0239935P.

PR 13-OCT-2000; 2000US-0239937P.

PR 20-OCT-2000; 2000US-0240960P.

PR 20-OCT-2000; 2000US-0241221P.

PR 20-OCT-2000; 2000US-0241785P.

PR 20-OCT-2000; 2000US-0241786P.

PR 20-OCT-2000; 2000US-0241787P.

PR 20-OCT-2000; 2000US-0241808P.

PR 20-OCT-2000; 2000US-0241809P.

PR 20-OCT-2000; 2000US-0241826P.

PR 01-NOV-2000; 2000US-0244617P.

PR 08-NOV-2000; 2000US-0246475P.

PR 08-NOV-2000; 2000US-0246476P.

PR 08-NOV-2000; 2000US-0246477P.

PR 08-NOV-2000; 2000US-0246478P.

PR 08-NOV-2000; 2000US-0246523P.

PR 08-NOV-2000; 2000US-0246524P.

PR 14-AUG-2000; 2000US-0225266P.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

CM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 23:23:40 ; Search time 45.2027 Seconds
(without alignments)
6113.919 Million cell updates/sec

Title: US-09-936-456-3
Perfect score: 498

Sequence: 1 cttcagcgtgcgcctcgcgc.....aaaaataaagtcgacc 498

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA.*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTCUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
C 1	21	4.2	623 3 US-09-385-982-207	Sequence 207, App
C 2	21	4.2	1492 4 US-09-976-594-937	Sequence 937, App
C 3	20	4.0	3765 3 US-07-705-490-1	Sequence 1, Appli
C 4	20	4.0	3765 3 US-07-751-891B-1	Sequence 1, Appli
C 5	20	4.0	4154 1 US-08-131-365B-37	Sequence 37, Appl
C 6	20	4.0	4154 2 US-08-668-123-37	Sequence 37, Appl
C 7	20	4.0	4362 2 US-08-453-073A-1	Sequence 1, Appli
C 8	20	4.0	6732 4 US-09-976-594-99	Sequence 99, Appl
C 9	20	4.0	640681 4 US-09-790-988-1	Sequence 1, Appli
C 10	19	3.8	400 4 US-08-956-171E-3987	Sequence 3987, Ap
C 11	19	3.8	4820 4 US-09-621-976-19147	Sequence 19147, A
C 12	19	3.8	522 4 US-09-621-976-18831	Sequence 18831, A
C 13	19	3.8	543 4 US-09-976-594-498	Sequence 498, App
C 14	19	3.8	546 4 US-09-220-123-151	Sequence 151, App
C 15	19	3.8	1101 4 US-09-486-382B-7	Sequence 7, Appli
C 16	19	3.8	1498 4 US-09-486-382B-3	Sequence 3, Appli
C 17	19	3.8	1843 4 US-09-807-258-3	Sequence 3, Appli
C 18	19	3.8	3410 4 US-09-016-434-1080	Sequence 1080, Ap
C 19	19	3.8	5229 4 US-09-635-872A-16	Sequence 16, Appl
C 20	19	3.8	5229 4 US-09-636-077A-16	Sequence 16, Appl
C 21	19	3.8	5229 4 US-09-636-067C-16	Sequence 16, Appl
C 22	19	3.8	5229 4 US-09-986-552-16	Sequence 16, Appl
C 23	19	3.8	6040 4 US-10-204-708-70	Sequence 70, Appl
C 24	19	3.8	7786 4 US-09-790-988-2	Sequence 2, Appli
C 25	19	3.8	8174 1 US-07-914-281-5	Sequence 5, Appli
C 26	19	3.8	8174 1 US-08-393-246-5	Sequence 5, Appli
C 27	19	3.8	8174 1 US-08-525-058A-5	Sequence 5, Appli

C 28	19	3.8	8174 2 US-08-696-731-5	Sequence 5, Appli
C 29	19	3.8	8174 3 US-09-042-531-5	Sequence 5, Appli
C 30	19	3.8	8174 5 PCT-US91-00899-3	Sequence 3, Appli
C 31	19	3.8	19650 4 US-09-819-989-3	Sequence 3, Appli
C 32	19	3.8	19650 4 US-10-373-992-3	Sequence 3, Appli
C 33	19	3.8	162450 4 US-09-345-882-1	Sequence 1, Appli
C 34	18	3.6	216 4 US-09-328-352-1507	Sequence 1507, Ap
C 35	18	3.6	401 4 US-09-621-976-11383	Sequence 11383, A
C 36	18	3.6	402 4 US-09-621-976-18325	Sequence 18325, A
C 37	18	3.6	528 4 US-09-621-976-171	Sequence 171, App
C 38	18	3.6	618 4 US-09-134-000C-2782	Sequence 709, App
C 39	18	3.6	618 4 US-09-134-000C-2782	Sequence 2782, Ap
C 40	18	3.6	851 4 US-09-008-892-10	Sequence 10, Appl
C 41	18	3.6	1081 4 US-09-976-594-522	Sequence 522, App
C 42	18	3.6	1143 4 US-09-252-991A-12947	Sequence 12947, A
C 43	18	3.6	1158 4 US-09-328-352-1515	Sequence 1515, Ap
C 44	18	3.6	1248 2 US-09-200-141-1	Sequence 1, Appli
C 45	18	3.6	1633 1 US-08-441-629-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-09-385-982-207/c
Sequence 207, Application US/09385982
Patent No. 6262334
GENERAL INFORMATION:
APPLICANT: ENDEGE, WILSON O., ET AL.
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCDNA-260XX
CURRENT APPLICATION NUMBER: US/09/385,982
CURRENT FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 09/328,111
EARLIER FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: 60/117,393
EARLIER FILING DATE: 1999-01-27
EARLIER APPLICATION NUMBER: 60/098,639
EARLIER FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 544
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 207
LENGTH: 623
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(623)
OTHER INFORMATION: n = A,T,C or G
US-09-385-982-207

Query Match 4.2% Score 21; DB 3; Length 623;
Best Local Similarity 100.0%; Pred. No. 0.81;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 TTTTGTGTTTTTAAAAAAT 487
DB 117 TTTTGTGTTTTTAAAAAAT 97

RESULT 2
US-09-976-594-937
Sequence 937, Application US/0976594
Patent No. 6673549
GENERAL INFORMATION:
APPLICANT: Furness, Michael
APPLICANT: Buchbinder, Jenny
TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STERC
FILE REFERENCE: PA-0041 US
CURRENT APPLICATION NUMBER: US/09/976,594
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: 60/240,409

```

; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 937
; LENGTH: 1492
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 3257507CB1
US-09-976-594-937

Query Match
Best Local Similarity 100.0%; Pred. No. 0.83;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 TTTTGTCTTTTAAAAAAAT 487
Db 1394 TTTGTCTTTTAAAAAAAT 1414

RESULT 3
US-07-705-490-1
; Sequence 1, Application US/07705490
; Patent No. 6107025
; GENERAL INFORMATION:
; APPLICANT: Caskey, C. T.
; APPLICANT: Nelson, David L.
; APPLICANT: Pieretti, Maura
; APPLICANT: Warren, Stephen T.
; APPLICANT: Oostira, Ben A.
; TITLE OF INVENTION: Diagnosis of the Fragile X Syndrome
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Thomas D. Paul
; STREET: 1301 McKinney, Suite 5100
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77010-3095
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/705,490
; FILING DATE: 19910708
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul, Thomas D.
; REGISTRATION NUMBER: 32,714
; REFERENCE/DOCKET NUMBER: D-5350
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713/651-5325
; TELEFAX: 713/651-5246
; TELEX: 762829
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3765 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-07-705-490-1

Query Match
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 458 CTTTCTCTCTCTCTCTTTT 477
Db 458 CTTTCTCTCTCTCTCTTTT 477
```

```

Db 3159 CTTTCTCTCTCTCTTTT 3178

RESULT 4
US-07-751-891B-1
; Sequence 1, Application US/07751891B
; Patent No. 6180337
; GENERAL INFORMATION:
; APPLICANT: Caskey, C. T.
; APPLICANT: Nelson, David L.
; APPLICANT: Pieretti, Maura
; APPLICANT: Warren, Stephen T.
; APPLICANT: Oostira, Ben A.
; APPLICANT: Fu, Ying-hui
; TITLE OF INVENTION: Diagnosis of the Fragile X Syndrome
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Thomas D. Paul
; STREET: 1301 McKinney, Suite 5100
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77010-3095
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/751,891B
; FILING DATE: 28-Aug-1991
; CLASSIFICATION: <unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul, Thomas D.
; REGISTRATION NUMBER: 32,714
; REFERENCE/DOCKET NUMBER: D-5350
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713/651-5325
; TELEFAX: 713/651-5246
; TELEX: 762829
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3765 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-07-751-891B-1

Query Match
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 458 CTTTCTCTCTCTCTTTT 477
Db 3159 CTTTCTCTCTCTCTTTT 3178

RESULT 5
US-08-131-365B-37
; Sequence 37, Application US/08131365B
; Patent No. 5527690
; GENERAL INFORMATION:
; APPLICANT: Brown, Michael S.
; APPLICANT: Briggs, Michael R.
; APPLICANT: Wang, Xiaodong
; APPLICANT: Goldstein, Joseph L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
; TO STEROID REGULATORY ELEMENT BINDING
```

TITLE OF INVENTION: PROTEINS
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: U.S.A.
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/131,365B
FILING DATE: 01-OCT-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: UTSD:372/PAR
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 4154 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
FEATURE:
NAME/KEY: CDS
LOCATION: 167..3607
US-08-131-365B-37

Query Match 4.0%; Score 20; DB 1; Length 4154;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 337 AGCAGTGGCCGCGCAGGCGCTG 356
Db 2393 AGCAGTGGCCGCGCAGGCGCTG 2412

RESULT 6
US-08-668-123-37
Sequence 37, Application US/08668123
Patent No. 5891631
GENERAL INFORMATION:
APPLICANT: Brown, Michael S.
APPLICANT: Briggs, Michael R.
APPLICANT: Wang, Xiaodong
APPLICANT: Goldstein, Joseph L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
TITLE OF INVENTION: TO STEROL REGULATORY ELEMENT BINDING
TITLE OF INVENTION: PROTEINS
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: U.S.A.
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/668,123
FILING DATE: 14-JUN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/131,365
FILING DATE: 01-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: UTSD:372/PAR
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 4154 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
FEATURE:
NAME/KEY: CDS
LOCATION: 167..3607
US-08-668-123-37

Query Match 4.0%; Score 20; DB 2; Length 4154;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 337 AGCAGTGGCCGCGCAGGCGCTG 356
Db 2393 AGCAGTGGCCGCGCAGGCGCTG 2412

RESULT 7
US-08-455-073A-1
Sequence 1, Application US/08455073A
Patent No. 5876949
GENERAL INFORMATION:
APPLICANT: Gideon Dreyfuss
APPLICANT: Mikiko C. Sioml
APPLICANT: Yan Zhang
TITLE OF INVENTION: Fragile X Related Proteins, Compositions And Methods
TITLE OF INVENTION: Of Making And Using The Same
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5876949ris
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/455,073A
FILING DATE: 31-MAY-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: UPN-2201
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4362 base pairs
TYPE: nucleic acid

```
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURES:
NAME/KEY: CDS
LOCATION: 220..2118
US-08-455-073A-1

Query Match
Best Local Similarity 100.0%; Score 20; DB 2; Length 4362;
Pred. No. 2.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 458 CTTTCTTTCTTTGTTTTT 477
Db 3303 CTTTCTTTCTTTGTTTTT 3322

RESULT 8
US-09-976-594-99
Sequence 99, Application US/09976594
Patent No. 6673549
GENERAL INFORMATION:
APPLICANT: Furness, Michael
APPLICANT: Buchbinder, Jenny
TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
FILE REFERENCE: PA-0041 US
CURRENT APPLICATION NUMBER: US/09/976,594
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: 60/240,409
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 1143
SOFTWARE: PERL Program
SEQ ID NO 99
LENGTH: 6732
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. 6673549 272843.14
LOCATION: 444, 4904, 4907, 4910, 4914
OTHER INFORMATION: a, t, c, g, or other
US-09-976-594-99

Query Match
Best Local Similarity 100.0%; Score 20; DB 4; Length 6732;
Pred. No. 2.7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 337 AGCAGTGGCCGCCAGGCTG 356
Db 2484 AGCAGTGGCCGCCAGGCTG 2503

RESULT 9
US-09-790-988-1
Sequence 1, Application US/09790988
Patent No. 6632935
GENERAL INFORMATION:
APPLICANT: SHIGEMOHU, SHUJI
APPLICANT: MATANABE, HIDEMI
APPLICANT: HATORI, MASAHIRA
APPLICANT: SAKAKI, YOSHIYUKI
TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
FILE REFERENCE: 081356/0159
CURRENT APPLICATION NUMBER: US/09/790,988
CURRENT FILING DATE: 2001-02-23
PRIOR APPLICATION NUMBER: JP2000-107160
PRIOR FILING DATE: 2000-04-07
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 640681
TYPE: DNA
```

```
ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match
Best Local Similarity 100.0%; Score 20; DB 4; Length 640681;
Pred. No. 3.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 TTTTAAAAAAATATAA 491
Db 591945 TTTTAAAAAAATATAA 591964

RESULT 10
US-08-956-171E-3987
Sequence 3987, Application US/08956171E
Patent No. 6593114
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
APPLICANT: Gil H. Choi
APPLICANT: Patrick S. Dillon
APPLICANT: Craig A. Rosen
APPLICANT: Steven C. Barash
APPLICANT: Michael R. Fannon
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5256
CORRESPONDENCE ADDRESS:
ADDRESS: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,171E
FILING DATE: 20-Oct-1997
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,861
FILING DATE: January 5, 1996
APPLICATION NUMBER: 08/781,986
FILING DATE: January 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Mark J. Hyman
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB248P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (240) 314-1224
FAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 3987:
SEQUENCE CHARACTERISTICS:
LENGTH: 400 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 3987:
US-08-956-171E-3987

Query Match
Best Local Similarity 100.0%; Score 19; DB 4; Length 400;
Pred. No. 7.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 238 AAGCAGGCGGCTCTCCAG 256
Db 332 AAGCAGGCGGCTCTCCAG 350

RESULT 11
US-09-621-976-19147
```

```
/ Sequence 19147, Application US/09621976
/ Patent No. 6639063
/ GENERAL INFORMATION:
/ APPLICANT: Dumas Milne Edwards, J.B.
/ APPLICANT: Jobert, S.
/ APPLICANT: Giordano, J.Y.
/ TITLE OF INVENTION: ESTs and Encoded Human Proteins.
/ FILE REFERENCE: GENSET.054PR2
/ CURRENT APPLICATION NUMBER: US/09/621,976
/ CURRENT FILING DATE: 2000-07-21
/ NUMBER OF SEQ ID NOS: 19335
/ SOFTWARE: Patent.pm
/ SEQ ID NO 19147
/ LENGTH: 480
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-621-976-19147
```

```
Query Match 3.8%; Score 19; DB 4; Length 480;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 459 TTTTCTTTCTTTGTTTTT 477
DB 22 TTTTCTTTCTTTGTTTTT 40
```

```
RESULT 12
US-09-621-976-18831/c
/ Sequence 18831, Application US/09621976
/ Patent No. 6639063
/ GENERAL INFORMATION:
/ APPLICANT: Dumas Milne Edwards, J.B.
/ APPLICANT: Jobert, S.
/ APPLICANT: Giordano, J.Y.
/ TITLE OF INVENTION: ESTs and Encoded Human Proteins.
/ FILE REFERENCE: GENSET.054PR2
/ CURRENT APPLICATION NUMBER: US/09/621,976
/ CURRENT FILING DATE: 2000-07-21
/ NUMBER OF SEQ ID NOS: 19335
/ SOFTWARE: Patent.pm
/ SEQ ID NO 18831
/ LENGTH: 522
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-621-976-18831
```

```
Query Match 3.8%; Score 19; DB 4; Length 522;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 468 TTTGTTTTTAAAAAAA 486
DB 100 TTTGTTTTTAAAAAAA 82
```

```
RESULT 13
US-09-976-594-498/c
/ Sequence 498, Application US/0976594
/ Patent No. 6673549
/ GENERAL INFORMATION:
/ APPLICANT: Furness, Michael
/ APPLICANT: Buchinder, Jenny
/ TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
/ FILE REFERENCE: PA-0041 US
/ CURRENT APPLICATION NUMBER: US/09/976,594
/ CURRENT FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: 60/240,409
/ PRIOR FILING DATE: 2000-10-12
/ NUMBER OF SEQ ID NOS: 1143
/ SOFTWARE: PERL Program
/ SEQ ID NO 498
/ LENGTH: 543
```

```
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc.feature
/ OTHER INFORMATION: incyte ID No. 6673549 223416.15
US-09-976-594-498
```

```
Query Match 3.8%; Score 19; DB 4; Length 543;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 459 TTTTCTTTCTTTGTTTTT 477
DB 351 TTTTCTTTCTTTGTTTTT 333
```

```
RESULT 14
US-09-220-132-151/c
/ Sequence 151, Application US/09220132
/ Patent No. 650607
/ GENERAL INFORMATION:
/ APPLICANT: Shyjan, Andrew W.
/ TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT
/ TITLE OF INVENTION: OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE C
/ FILE REFERENCE: 07334-074001
/ CURRENT APPLICATION NUMBER: US/09/220,132
/ CURRENT FILING DATE: 1998-12-23
/ PRIOR APPLICATION NUMBER: US 60/079,303
/ PRIOR FILING DATE: 1998-03-25
/ PRIOR APPLICATION NUMBER: US 60/068,821
/ PRIOR FILING DATE: 1997-12-24
/ NUMBER OF SEQ ID NOS: 191
/ SOFTWARE: PaetSeq for Windows Version 4.0
/ SEQ ID NO 151
/ LENGTH: 546
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc.feature
/ LOCATION: (1)...(546)
/ OTHER INFORMATION: n = A,T,C or G
US-09-220-132-151
```

```
Query Match 3.8%; Score 19; DB 4; Length 546;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 468 TTTGTTTTTAAAAAAA 486
DB 25 TTTGTTTTTAAAAAAA 7
```

```
RESULT 15
US-09-486-382B-7
/ Sequence 7, Application US/09486382B
/ Patent No. 6388174
/ GENERAL INFORMATION:
/ APPLICANT: Hokko Chemical Industry Co., Ltd.
/ APPLICANT: MINISTRY OF AGRICULTURE, FORESTRY AND FISHERIES
/ TITLE OF INVENTION: Genes for encoding alpha-subunits of anthranilate
/ TITLE OF INVENTION: synthase of rice, and DNAs related to said genes
/ FILE REFERENCE: 10647
/ CURRENT APPLICATION NUMBER: US/09/486,382B
/ CURRENT FILING DATE: 2001-06-11
/ PRIOR APPLICATION NUMBER: JP 9-235049
/ PRIOR FILING DATE: 1997-08-29
/ NUMBER OF SEQ ID NOS: 19
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 7
/ LENGTH: 1101
/ TYPE: DNA
/ ORGANISM: Oryza sativa
/ FEATURE:
```

NAME/KEY: promoter
LOCATION: (1)..(1101)
US-09-486-382B-7

Query Match 3.8%; Score 19; DB 4; Length 1101;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 467 TTTGTTTTTAAAAAA 485
Db 51 TTTGTTTTTAAAAAA 69

Search completed: February 24, 2004, 03:24:18
Job time : 47.2027 secs


```
QY 242 AGCGCGCTCTCCAGACTTAATATATACCACTAACCTTGAGGGGGAGCCCAATCTGG 301
DB 1204 AGCGCGCTCTCCAGACTTAATATATATACCACTAACCTTGAGGGGGAGCCCAATCTGG 1263
QY 302 ACTCCCTCCCGGCTTTGGGACATCGCAGGCGGGAGAGAGAGCCCGCAGGCTGGGGCCA 361
DB 1264 ACTCCCTCCCGGCTTTGGGACATCGCAGGCGGGAGAGAGAGAGCCCGCAGGCTGGGGCCA 1323
QY 362 GGAAGAGCTCCAGAAAGGAGCACTGAGCGCTGTGGCGGAGGCTTCGACATCCGAGGCA 421
DB 1324 GGAAGAGCTCCAGAAAGGAGCACTGAGCGCTGTGGCGGAGGCTTCGACATCCGAGGCA 1383
QY 422 CCAGGGAAGTCTCCTGGGGGAGATCTGTAAATAAAGCTTTTCTTTTCTTTTAA 481
DB 1384 CCAGGGAAGTCTCCTGGGGGAGATCTGTAAATAAAGCTTTTCTTTTCTTTTAA 1443
QY 482 AAAAA 486
DB 1444 AAAAA 1448
```

```
RESULT 2
US-10-187-657-2
; Sequence 2, Application US/10187657
; Publication No. US2003006831A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy K.W.
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Azimzal, Yalda
; TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
; FILE REFERENCE: PV-0009 CIP
; CURRENT APPLICATION NUMBER: US/10/187,657
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: PCT/US00/07817
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/139,565
; PRIOR FILING DATE: 1999-06-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program
; SEQ ID NO 2
; LENGTH: 1421
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US2003006831A1 4901066CB1
US-10-187-657-2
```

```
Query Match 83.5%; Score 416; DB 14; Length 1421;
Best Local Similarity 100.0%; Pred. No. 2.8e-199;
Matches 416; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 62 GCGCGCGCTTGGGTCACGCTGCAACCGGCGTCTGTGCTCTTCTCGAGGGGCG 121
DB 1006 GCGCGCGCTTGGGTCACGCTGCAACCGGCGTCTGTGCTCTTCTCGAGGGGCG 1065
QY 122 TGTGAGTCTCCAGATATGTTGGCCCGAGGCTCTTGCACCTTTGTGAACCAAGCGCA 181
DB 1066 TGTGAGTCTCCAGATATGTTGGCCCGAGGCTCTTGCACCTTTGTGAACCAAGCGCA 1125
QY 182 AGGACTGCAGCCAGAGAGAGGGGCTCACTTTATCCTCGGCAACCACTGCACAGC 241
DB 1126 AGGACTGCAGCCAGAGAGAGGGGCTCACTTTATCCTCGGCAACCACTGCACAGC 1185
QY 242 AGGCGGCTCTCCAGACTTAATAATGATACCACTTAAGTGTAGAGGGAGCCCAATCTGG 301
DB 1186 AGGCGGCTCTCCAGACTTAATAATGATACCACTTAAGTGTAGAGGGAGCCCAATCTGG 1245
QY 302 ACTCCTTCCCGGCTTTGGGACATCGAGGCGGGAGAGCAAGTCCCGCAGGCTTGGGCA 361
DB 1246 ACTCCTTCCCGGCTTTGGGACATCGAGGCGGGAGAGCAAGTCCCGCAGGCTTGGGCA 1305
QY 362 GGAAGAGCTCCAGAAAGGAGCACTGAGCGCTGTGGCGGAGGCTTCGACATCCGAGGCA 421
```

```
DB 1306 GGAAGAGCTCCAGAAAGGAGCACTGAGCGCTCTGCGCGAGGCTTCGACATCCGAGGCA 1365
QY 422 CCAGGGAAGTCTCCTGGGGGAGATCTGTAAATAAAGCTTTTCTTTTCTTTT 477
DB 1366 CCAGGGAAGTCTCCTGGGGGAGATCTGTAAATAAAGCTTTTCTTTTCTTTT 1421
```

```
RESULT 3
US-10-187-657-4
; Sequence 4, Application US/10187657
; Publication No. US2003006831A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy K.W.
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Azimzal, Yalda
; TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
; FILE REFERENCE: PV-0009 CIP
; CURRENT APPLICATION NUMBER: US/10/187,657
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: PCT/US00/07817
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/139,565
; PRIOR FILING DATE: 1999-06-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program
; SEQ ID NO 4
; LENGTH: 1420
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US2003006831A1 322161CA2
US-10-187-657-4
```

```
Query Match 81.7%; Score 407; DB 14; Length 1420;
Best Local Similarity 100.0%; Pred. No. 9.5e-195;
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 62 GCGCGCGCTTGGGTCACGCTGCAACCGGCGTCTGTGCTCTTCTCGAGGGGCG 121
DB 1006 GCGCGCGCTTGGGTCACGCTGCAACCGGCGTCTGTGCTCTTCTCGAGGGGCG 1065
QY 122 TGTGAGTCTCCAGATATGTTGGCCCGAGGCTCTTGCACCTTTGTGAACCAAGCGCA 181
DB 1066 TGTGAGTCTCCAGATATGTTGGCCCGAGGCTCTTGCACCTTTGTGAACCAAGCGCA 1125
QY 182 AGGACTGCAGCCAGAGAGAGGGGCTCACTTTATCCTCGGCAACCACTGCACAGC 241
DB 1126 AGGACTGCAGCCAGAGAGAGGGGCTCACTTTATCCTCGGCAACCACTGCACAGC 1185
QY 242 AGGCGGCTCTCCAGACTTAATAATGATACCACTTAAGTGTAGAGGGAGCCCAATCTGG 301
DB 1186 AGGCGGCTCTCCAGACTTAATAATGATACCACTTAAGTGTAGAGGGAGCCCAATCTGG 1245
QY 302 ACTCCTTCCCGGCTTTGGGACATCGAGGCGGGAGAGCAAGTCCCGCAGGCTTGGGCA 361
DB 1246 ACTCCTTCCCGGCTTTGGGACATCGAGGCGGGAGAGCAAGTCCCGCAGGCTTGGGCA 1305
QY 362 GGAAGAGCTCCAGAAAGGAGCACTGAGCGCTGTGGCGGAGGCTTCGACATCCGAGGCA 421
DB 1306 GGAAGAGCTCCAGAAAGGAGCACTGAGCGCTGTGGCGGAGGCTTCGACATCCGAGGCA 1365
QY 422 CCAGGGAAGTCTCCTGGGGGAGATCTGTAAATAAAGCTTTTCTTTTCTTT 468
DB 1366 CCAGGGAAGTCTCCTGGGGGAGATCTGTAAATAAAGCTTTTCTTTTCTTT 1412
```

```
RESULT 4
US-10-094-749-1195/c
; Sequence 1195, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
```



```

APPLICANT: ISOGAI, TAKAO
APPLICANT: SUGIYAMA, TOMOYASU
APPLICANT: OTSUKI, TETSUJI
APPLICANT: WAKAMATSU, AI
APPLICANT: SATO, HIROYUKI
APPLICANT: ISHII, SHIZUKO
APPLICANT: YAMAMOTO, JUN-ICHI
APPLICANT: ISONO, YUUKO
APPLICANT: HTO, YURI
APPLICANT: OTSUKA, KAORU
APPLICANT: NAGAI, KEIICHI
APPLICANT: IRIE, RYOTARO
APPLICANT: TAMECHIKA, ICHIRO
APPLICANT: SEKI, NAOHICO
APPLICANT: YOSHIKAWA, TSUTOMU
APPLICANT: OTSUKA, MOTOMYUKI
APPLICANT: NAGAHARI, KENJI
APPLICANT: MASUHO, YASUHIKO
TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
FILE REFERENCE: 084335/0160
CURRENT APPLICATION NUMBER: US/10/094,749
CURRENT FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 60/350,435
PRIOR FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: JP 2001-328381
PRIOR FILING DATE: 2001-09-14
NUMBER OF SEQ ID NOS: 381
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1195
LENGTH: 2684
TYPE: DNA
ORGANISM: Homo sapiens
US-10-094-749-1195

```

Query	Match	Similarity	Score	DB	Length
Best Local Similarity	100.0%	Pred. No. 3,4e-187			
Matches 392	Conservative 0	Mismatches 0	Indels 0	Gaps 0	
QY	91	GAGCCTCTGTGCTCTTCTCTCGAGGGGCGGTGTGAATCTCCAGTATGTTCGGCCACG	150		
Db	2143	GAGGTCTGTGCTCTTCTCTCGAGGGGCGGTGTGAATCTCCAGTATGTTCGGCCACG	2284		
QY	151	GCTCTTCGACCCCTTCTGACCCAAGCGCGAAGCTCGACCCGAGAGAGGGGGCTCA	210		
Db	2283	GCTCTTCGACCCCTTCTGACCCAAGCGCGAAGCTCGACCCGAGAGAGGGGGCTCA	222		
QY	211	CCTTTATCTCGGCGACCACTGCACAAAGCGCGCTCCGACATTAAATGTATC	270		
Db	2223	CTCTTATCTCGGCGACCACTGCACAAAGCGCGCTCCGACATTAAATGTATC	216		
QY	271	ACCACATCACTGTGTAGGGGAGCCCAATCTGGAATCTTCCCGGCTTGGACATTCGAGG	330		
Db	2163	ACCACATCACTGTGTAGGGGAGCCCAATCTGGAATCTTCCCGGCTTGGACATTCGAGG	210		
QY	331	CCGGGAAGCAGTCCCGGCGAGGCTGTGGGCGAGAGAGCTCCAGGAAGGCGACATGAGCGCT	390		
Db	2103	CCGGGAAGCAGTCCCGGCGAGGCTGTGGGCGAGAGAGCTCCAGGAAGGCGACATGAGCGCT	204		
QY	391	GCTGGCGGAGGCTCTGGAATCTCGCAGGCGACAGAGGAAATCTCTGTGGGCGATCTGTA	450		
Db	2043	GCTGGCGGAGGCTCTGGAATCTCGCAGGCGACAGAGGAAATCTCTGTGGGCGATCTGTA	198		
QY	451	AATTAACCTTTTCTTTTGTTTTTTAAAAA	482		
Db	1983	AATTAACCTTTTCTTTTGTTTTTTAAAAA	1952		

RESULT 5
US-10-187-657-3
; Sequence 3, Application US/10187657
; Publication No. US20030068311A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy K.W.

```

1  APPLICANT: Baughn, Mariah R.
2  APPLICANT: Azimzai, Yalda
3  TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
4  FILE REFERENCE: PV-0009 CIP
5  CURRENT APPLICATION NUMBER: US/10/187,657
6  CURRENT FILING DATE: 2002-07-01
7  PRIOR APPLICATION NUMBER: PCT/US00/07817
8  PRIOR FILING DATE: 2000-03-22
9  PRIOR APPLICATION NUMBER: 60/133,565
10 PRIOR FILING DATE: 1999-06-16
11 NUMBER OF SEQ ID NOS: 9
12 SOFTWARE: PERL Program
13 SEQ ID NO 3
14 LENGTH: 346
15 TYPE: DNA
16 ORGANISM: Homo sapiens
17 FEATURE:
18 NAME/KEY: misc_feature
19 OTHER INFORMATION: Incyte ID No. US2003006831A1 1752794F6
20 FEATURE:
21 NAME/KEY: unsure
22 LOCATION: 91, 188, 206, 291, 337
23 OTHER INFORMATION: a, t, c, g, or other
24 US-10-187-657-3

```

Qy	82	CTGGCAACCGGGGCTCTGTGCTCTCTCTCTCTGGAGGGGGCGTGGTAGTCCAGTAGT	141
Db	92	CTGGCAACCGGGGCTCTGTGCTCTCTCTCTCTGGAGGGGGCGTGGTAGTCCAGTAGT	151
Qy	142	CGGCCAGCGCTCTTCTGGACCTTTCTGGACCAAG	177
Db	152	CGGCCAGCGCTCTTCTGGACCTTTCTGGACCAAG	187

Query Match 19.3%; Score 96; DB 14; Length 346;
 Best Local Similarity 100.0%; Pred. No. 4.3e-38;
 Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

RESULT 6
US-10-027-632-91970
; Sequence 91970, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 91970
; LENGTH: 439
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(439)
; OTHER INFORMATION: n = A,T,C or G

```

US-10-027-632-91970

Query Match 7.0%; Score 35; DB 15; Length 439;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 CCGCTAGGCTCCTCCGCGCTCACCCTCAGTACG 60
DB 265 CCGCTAGGCTCCTCCGCGCTCACCCTCAGTACG 319

RESULT 7
US-10-424-599-78751/c
Sequence 78751, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:

APPLICANT: La Rosa, Thomas J
APPLICANT: Kovalic, David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 78751
LENGTH: 408
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(408)
OTHER INFORMATION: unsure at all n locations
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_42127C.1
US-10-424-599-78751

Query Match 4.6%; Score 23; DB 12; Length 408;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 469 TTGTTTTTTTAAAAAATATATAA 491
DB 86 TTGTTTTTTTAAAAAATATATAA 64

RESULT 8
US-10-027-632-172855/c
Sequence 172855, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:

APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 172855
LENGTH: 822
TYPE: DNA
ORGANISM: Human
US-10-027-632-172855

Query Match 4.4%; Score 22; DB 15; Length 822;
Best Local Similarity 100.0%; Pred. No. 0.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 459 TTTTTCCTTTTGTGTTTTTAAA 480
DB 492 TTTTTCCTTTTGTGTTTTTAAA 471

RESULT 9
US-10-027-632-172856/c
Sequence 172856, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:

APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 172856
LENGTH: 822
TYPE: DNA
ORGANISM: Human
US-10-027-632-172856

Query Match 4.4%; Score 22; DB 15; Length 822;
Best Local Similarity 100.0%; Pred. No. 0.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 459 TTTTTCCTTTTGTGTTTTTAAA 480
DB 492 TTTTTCCTTTTGTGTTTTTAAA 471

RESULT 10
US-10-027-632-29944
Sequence 29944, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20

```

; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 29944
; LENGTH: 911
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-29944
```

```

Query Match
Best Local Similarity 4.4%; Score 22; DB 15; Length 911;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 450 AATTAACCTTTTCTTTG 471
DB 612 AATTAACCTTTTCTTTG 633
```

```

RESULT 11
US-10-260-238-2487
; Sequence 2487, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Krebs, Joel
; APPLICANT: Provatt, Nicholas
; APPLICANT: Rieke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO: 2487
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Oryza sativa
US-10-260-238-2487
```

```

Query Match
Best Local Similarity 4.4%; Score 22; DB 15; Length 2000;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 472 TTTTAAATTAATTAAGT 493
DB 843 TTTTAAATTAATTAAGT 864
```

```

RESULT 12
US-09-764-847-1176/C
; Sequence 1176, Application US/09764847
; Patent No. US20020132767A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO: 1176
; LENGTH: 3243
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-847-1176
```

```

Query Match
Best Local Similarity 4.4%; Score 22; DB 9; Length 3243;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 370 CCAGGAAGGCGACTGAGCGCTG 391
DB 2584 CCAGGAAGGCGACTGAGCGCTG 2563
```

```

RESULT 13
US-10-092-154-1176/C
; Sequence 1176, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009C1
; CURRENT APPLICATION NUMBER: US/10/092,154
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2003
; Prior Application removed - See file wrapper or Palm
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO: 1176
; LENGTH: 3243
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-092-154-1176
```

```

Query Match
Best Local Similarity 4.4%; Score 22; DB 14; Length 3243;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 370 CCAGGAAGGCGACTGAGCGCTG 391
DB 2584 CCAGGAAGGCGACTGAGCGCTG 2563
```

```

RESULT 14
US-09-960-352-7720/C
; Sequence 7720, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagesan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO: 7720
; LENGTH: 325
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 33-LIB34-009-Q1-E1-A2
US-09-960-352-7720
```

Query Match 4.2%; Score 21; DB 9; Length 325;
 Best Local Similarity 100.0%; Pred. No. 2.6;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 CCTTTTTCCTTTGTTTTT 477
 |||||
 DB 320 CCTTTTTCCTTTGTTTTT 300

RESULT 15

US-10-424-599-47139/C
 ; Sequence 47139; Application US/10424589
 ; Publication No. US20040031072A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa Thomas J
 ; APPLICANT: Kovalic David K
 ; APPLICANT: Zhou Yihua
 ; APPLICANT: Cao Yongwei
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53223) B
 ; CURRENT APPLICATION NUMBER: US/10/424,599
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 285684
 ; SEQ ID NO 47139
 ; LENGTH: 372
 ; TYPE: DNA
 ; ORGANISM: Glycine max
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT3847_142572C.1
 US-10-424-599-47139

Query Match 4.2%; Score 21; DB 12; Length 372;
 Best Local Similarity 100.0%; Pred. No. 2.6;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 459 TTTTTCCTTTGTTTTTA 479
 |||||
 DB 353 TTTTTCCTTTGTTTTTA 333

Search completed: February 24, 2004, 03:37:00
 Job time : 193.827 secs

```
Location/Qualifiers
1. .446
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
```


ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 434)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert length: 558 Std Error: 0.00
Seq primer: -400P from Gibco.

FEATURES

Source

Location/Qualifiers
1..434
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2522647"
/tissue_type="colonic mucosa from 3 patients with Crohn's
disease"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Soares Dieckgraefe Colon NHCD"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAAGTGGAGGCGCGCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Tissue samples
provided by Dr. Brian Dieckgraefe (Washington University,
dieck@um.wustl.edu); colonic mucosa represents a range of
disease involvement from moderate to severe Crohn's
disease; samples include both perforating (Istulas) and
non-perforating samples. Library constructed by Bento
Soares and M. Fatima Bonaldo."

ORIGIN

Query Match

Best Local Similarity 83.7%; Score 417; DB 9; Length 434;
Best Local Similarity 100.0%; Pred. No. 3.2e-163;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 62 GCGCGCGCTTCTGGGTCACGCTGGCAACGCGCTCTCTGCTCTTCTCGAGAGGCGCG 121
Db 417 GCGCGCGCTTCTGGGTCACGCTGGCAACGCGCTCTCTGCTCTTCTCGAGAGGCGCG 358
QY 122 TGTGAGTCTCCAGTATGTTGGGCCAGCGCTCTTGGACCTCTTGGACCAAGGCCCA 181
Db 357 TGTGAGTCTCCAGTATGTTGGGCCAGCGCTCTTGGACCTCTTGGACCAAGGCCCA 298
QY 182 AGGACTGAGCAGAGAGAGAGGCGGCTCACCTCTTATCTCTGCGACCACTGCAAGC 241
Db 297 AGGACTGAGCAGAGAGAGAGGCGGCTCACCTCTTATCTCTGCGACCACTGCAAGC 238
QY 242 AGGCGGCTCTCCAGACTTAAATGTATCACTAACTGTGAGGCGGAGCCCAATCTGG 301
Db 237 AGGCGGCTCTCCAGACTTAAATGTATCACTAACTGTGAGGCGGAGCCCAATCTGG 178
QY 302 ACTGCTTCCCGGCTTGGGACATGCGAGGCGGAGAGCACTGCGCGCAGGCTGGGCGCA 361
Db 177 ACTGCTTCCCGGCTTGGGACATGCGAGGCGGAGAGCACTGCGCGCAGGCTGGGCGCA 118
QY 362 GAGAGGCTCCAGAGAGGCACTGAGCGCTGCTGGCGGAGGCTTCGAGCATCCGACAGCA 421
Db 117 GAGAGGCTCCAGAGAGGCACTGAGCGCTGCTGGCGGAGGCTTCGAGCATCCGACAGCA 58
QY 422 CCAAGGAAAGTCTCTTGGGCGATCTGTAATAAATCTTTTCTTTTCTTTTCTTTT 478
Db 57 CCAAGGAAAGTCTCTTGGGCGATCTGTAATAAATCTTTTCTTTTCTTTTCTTTT 1

```

RESULT 4

AI304327 510 bp mRNA linear EST 01-FEB-1999
LOCUS 9057906.x1 NCI CGAP C08 Homo sapiens cDNA clone IMAGE:1912667 3'
DEFINITION similar to contains TARI.13 TARI repetitive element ;, mRNA
sequence.

ACCESSION

AI304327
VERSION AI304327.1 GI:3988016
KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 510)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

REFERENCE

AUTHORS

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov

Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnl.gov/bdnp/image/image.html
Insert length: 1467 Std Error: 0.00
Seq primer: -400P from Gibco

High quality sequence stop: 476.
Location/Qualifiers
1..510
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1912667"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: Colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."

FEATURES

Source

Location/Qualifiers
1..510
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1912667"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: Colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match

Best Local Similarity 83.5%; Score 416; DB 9; Length 510;
Best Local Similarity 100.0%; Pred. No. 7.7e-163;
Matches 416; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 62 GCGCGCGCTTCTGGGTCACGCTGGCAACGCGGCTCTGCTCTTCTCGAGAGGCGCG 121
Db 416 GCGCGCGCTTCTGGGTCACGCTGGCAACGCGGCTCTGCTCTTCTCGAGAGGCGCG 357
QY 122 TGTGAGTCTCCAGTATGTTGGGCCAGCGCTCTTGGACCTCTTGGACCAAGGCCCA 181
Db 356 TGTGAGTCTCCAGTATGTTGGGCCAGCGCTCTTGGACCTCTTGGACCAAGGCCCA 297
QY 182 AGGACTGAGCAGAGAGAGAGGCGGCTCACCTCTTATCTCTGCGACCACTGCAAGC 241
Db 296 AGGACTGAGCAGAGAGAGAGGCGGCTCACCTCTTATCTCTGCGACCACTGCAAGC 237
QY 242 AGGCGGCTCTCCAGACTTAAATGTATCACTAACTGTGAGGCGGAGCCCAATCTGG 301
Db 236 AGGCGGCTCTCCAGACTTAAATGTATCACTAACTGTGAGGCGGAGCCCAATCTGG 177
QY 302 ACTGCTTCCCGGCTTGGGACATGCGAGGCGGAGAGCACTGCGCGCAGGCTGGGCGCA 361
Db 176 ACTGCTTCCCGGCTTGGGACATGCGAGGCGGAGAGCACTGCGCGCAGGCTGGGCGCA 117

```

QY 362 GGAGAGCTCCAGAAAGGCACTGAGCGCTGCTGGCGAGGCGCTCGACATCGCAGCA 421
 Db 116 GGAGAGCTCCAGAAAGGCACTGAGCGCTGCTGGCGAGGCGCTCGACATCGCAGCA 57
 QY 422 CCAGGAAAGTCTCTCTGGGCGCATCTGTAAATAAACCTTTTCTTTTGT 477
 Db 56 CCAGGAAAGTCTCTCTGGGCGCATCTGTAAATAAACCTTTTCTTTTGT 1

RESULT 5
 AA593860/c 414 bp mRNA linear EST 25-SEP-1997
 LOCUS n19f03.s1 NCI_CGAP_Co12 Homo sapiens cDNA clone IMAGE:1084349 3'
 DEFINITION mRNA sequence.
 ACCESSION AA593860
 VERSION AA593860
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 414)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 Tumor Gene Index
 Unpublished (1997)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgabbs-remail.nih.gov
 Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Stratagene, Inc.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/bbrp/image/image.html
 Insert Length: 1204 Std Error: 0.00
 Seg primer: -40m3 fwd. RT from Amersham
 High quality sequence stop: 40c.
 Location/Qualifiers
 1..414
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1084349"
 /sex="mixed"
 /tissue_type="colon tumor"
 /lab_host="SOLR (kanamycin resistant)"
 /clone_lib="NCI CGAP Co12"
 /note="Organ: colon; Vector: Bluescript SK-; Site: 1;
 EcorI; Site 2: XhoI; Cloned unidirectionally. Primer:
 oligo dt. Pooled colon tumors. 5' adaptor sequence: 5'
 GAATTCGGCAGCAG 3' 3' adaptor sequence: 5'
 CTCGAGTTTCTTTTCTTTTCTTTT 3' Average insert size: 1.2 kb."

ORIGIN
 Query Match. 83.1%; Score 414; DB 9; Length 414;
 Best Local Similarity 100.0%; Pred. No. 5.8e-162;
 Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 CGCGGCTTCTGGGTACGCTGGCAACCGGCTCTCTGCTCTCTCGAGGCGCGCT 122
 Db 414 CGCGGCTTCTGGGTACGCTGGCAACCGGCTCTCTGCTCTCTCGAGGCGCGCT 355
 QY 123 GGTGAGTCTCAATATGTTGGGCCAGCGCTTTCGACCTTTCGACCAAGGCCAA 182
 Db 354 GGTGAGTCTCAATATGTTGGGCCAGCGCTTTCGACCTTTCGACCAAGGCCAA 295
 QY 183 GGTGAGTCTCAATATGTTGGGCCAGCGCTTTCGACCTTTCGACCAAGGCCAA 242
 Db 294 GGTGAGTCTCAATATGTTGGGCCAGCGCTTTCGACCTTTCGACCAAGGCCAA 235

QY 243 GCGCGCTCTCCAGACTTAAATATATATCACTAACCTGTGAGGGGACCAATCTGA 302
 Db 234 GCGCGCTCTCCAGACTTAAATATATATCACTAACCTGTGAGGGGACCAATCTGA 175
 QY 303 CTCTTCTCCCGCTTGGGACATCGCAGGCGCGGAGAGAGTCCCGCCAGGCTGGGCGAG 362
 Db 174 CTCTTCTCCCGCTTGGGACATCGCAGGCGCGGAGAGAGTCCCGCCAGGCTGGGCGAG 115
 QY 363 GAGAGCTCCAGAAAGGCACTGAGCGCTGCTGGCGAGGCGCTCGACATCGCAGGCA 422
 Db 114 GAGAGCTCCAGAAAGGCACTGAGCGCTGCTGGCGAGGCGCTCGACATCGCAGGCA 55
 QY 423 CCAGGAAAGTCTCTCTGGGCGCATCTGTAAATAAACCTTTTCTTTTGT 476
 Db 54 CCAGGAAAGTCTCTCTGGGCGCATCTGTAAATAAACCTTTTCTTTTGT 1

RESULT 6
 A1821606/c 537 bp mRNA linear EST 13-DEC-1999
 LOCUS nk08all.x5 NCI CGAP Co2 Homo sapiens cDNA clone IMAGE:1012892 3'
 DEFINITION similar to contains TAR1.c3 TAR1 MER22 repetitive element ; , mRNA
 sequence.
 ACCESSION A1821606
 VERSION A1821606.1 GI:5440685
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 537)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 Tumor Gene Index
 Unpublished (1997)
 JOURNAL Other ESTs: nk08all.y5
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Stratagene, Inc., David B. Krizman,
 Ph.D.
 CDNA Library Arraying: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/bbrp/image/image.html

This read is a RESEQUENCE of a previously sequenced human clone
 Original clone citation: National Cancer Institute, Cancer Genome
 Anatomy Project (CGAP), Tumor Gene Index.
 This read has been verified (found to hit its original self in the
 correct orientation).
 Insert Length: 741 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 459.
 Location/Qualifiers
 1..537
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1012892"
 /tissue_type="tumor"
 /lab_host="SOLR (kanamycin resistant)"
 /clone_lib="NCI CGAP Co2"
 /note="Organ: colon; Vector: Bluescript SK-; Site: 1;
 EcorI; Site 2: XhoI; Cloned unidirectionally. Primer:
 oligo dt. Bulk colon villosus adenoma. 5' adaptor sequence:
 5' GAATTCGGCAGCAG 3' 3' adaptor sequence: 5'
 CTCGAGTTTCTTTTCTTTTCTTTT 3' Average insert size: 1.1 kb."

ORIGIN
 Query Match 81.9%; Score 408; DB 9; Length 537;

Best Local Similarity 100.0%; Pred.No.1,6e-159;
Matches 408; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 62 GCGCGGCTTCTGGGTACAGCTGGCAACCGGCTCTTGTGCTTCTTCTTCTGAGAGGCGCG 121
Db 417 GCGCGGCTTCTGGGTACAGCTGGCAACCGGCTCTTGTGCTTCTTCTTCTGAGAGGCGCG 358
OY 122 TGTGAGTCTCCAGTATGTTTCGCGCCAGCGCTCTTTCGACCCCTTTCGACCAAGGCGCA 181
Db 357 TGTGAGTCTCCAGTATGTTTCGCGCCAGCGCTCTTTCGACCCCTTTCGACCAAGGCGCA 298
OY 182 AGAGCTGACGACGAG 241
Db 297 AGAGCTGACGACGAG 238
OY 242 AGCGCGCTCTCCAGACTTAATATGATACCACTAAGCTGTAGAGAGAGAGAGAGAGAGAG 301
Db 237 AGCGCGCTCTCCAGACTTAATATGATACCACTAAGCTGTAGAGAGAGAGAGAGAGAGAG 178
OY 302 ACTCTTCTCCGCTTGGGACATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 361
Db 177 ACTCTTCTCCGCTTGGGACATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 118
OY 362 GAGAGCTCCAG 421
Db 117 GAGAGCTCCAG 58
OY 422 CCAGAGAAAGTCTCTGAGGAGATCTTAATTAACCTTTTCTTT 469
Db 57 CCAGAGAAAGTCTCTGAGGAGATCTTAATTAACCTTTTCTTT 10

RESULT 7
A1274929/c 405 bp mRNA linear EST 29-JAN-1999
LOCUS q149c11.x1 NCI CGAP C08 Homo sapiens cDNA clone IMAGE:1875668 3'
DEFINITION similar to contains TAR1.t3 TAR1 repetitive element ;, mRNA
sequence.

ACCESSION A1274929
VERSION A1274929.1 GI:3897203
KEYWORDS EST.

ORGANISM Homo sapiens (human)

REFERENCE 1 (bases 1 to 405)
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov

ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 538)
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov

ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 538)
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov

ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

ORIGIN

Query Match 81.3%; Score 405; DB 9; Length 405;

Best Local Similarity 100.0%; Pred.No.3,3e-158;
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 73 TGGGTACAGCTGACCAACCGGCTCTGTCCTCTTCTTCGAGAGAGAGAGAGAGAGAGAG 132
Db 405 TGGGTACAGCTGACCAACCGGCTCTGTCCTCTTCTTCGAGAGAGAGAGAGAGAGAGAG 346
OY 133 CAGTATGTTGGGCGGCGGCTCTTTCGACACCTTTCGACCAAGGCGGCGGCGGCGGCGG 192
Db 345 CAGTATGTTGGGCGGCGGCTCTTTCGACACCTTTCGACCAAGGCGGCGGCGGCGGCGG 286
OY 193 CAG 252
Db 285 CAG 226
OY 253 CCAAGCTTAATATGATACCACTAAGCTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAG 312
Db 225 CCAAGCTTAATATGATACCACTAAGCTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 166
OY 313 GCGTGGGACATGACAG 372
Db 165 GCGTGGGACATGACAG 106
OY 373 GGAAGGACACTGAGCGGCTCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 432
Db 105 GGAAGGACACTGAGCGGCTCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 46
OY 433 CTCCTGGGCGGATCTTAATTAACCTTTTCTTTCTTTCTTTT 477
Db 45 CTCCTGGGCGGATCTTAATTAACCTTTTCTTTCTTTCTTTT 1

RESULT 8
A1660493/c 538 bp mRNA linear EST 18-DEC-1999
LOCUS w667h02.x1 Soares,Dieckgraefe,colon_MHCD Homo sapiens cDNA clone
DEFINITION IMAGE:2346195 3' similar to contains TAR1.t3 TAR1 repetitive
element ;, mRNA sequence.

ACCESSION A1660493
VERSION A1660493.1 GI:4764063
KEYWORDS EST.

ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 538)
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov

ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 538)
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov

ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 538)
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE
AUTHORS
TITLE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 381)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)

JOURNAL
COMMENT

Contact: Robert Strauberg, Ph.D.
Email: rgs@dsb-remail.nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bdnp/image/image.html
Insert Length: 1422 Std Error: 0.00
Seq primer: -40UP from Glibco
High quality sequence stop: 340.
Location/Qualifiers

FEATURES
source

1..381
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1873190"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/clone_lib="NCI CGAP Cos8"
/note="Organ: colon; Vector: pT773-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from a colon adenocarcinoma, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT773 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Ronaldo."

ORIGIN

Query Match 76.5%; Score 381; DB 9; Length 381;
Best Local Similarity 100.0%; Pred. No. 3,4e-148;

Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 97 CTGAGCTCTCTCTCGAGAGGGCGGTGTGAGTCTCCAGATATTTGGCCCAAGCGCTTT 156
DB 381 CTGAGCTCTCTCTCTCGAGAGGGCGGTGTGAGTCTCCAGATATTTGGCCCAAGCGCTTT 322
QY 157 CGACACCTTCTGACCAAGCGCCAGAGACTGACGCCAGAGAGAGAGGGGCTTACCTTT 216
DB 321 CGACACCTTCTGACCAAGCGCCAGAGACTGACGCCAGAGAGAGAGGGGCTTACCTTT 262
QY 217 ATCTCGGGGCGACCAAGCAAGAGAGGGCGCTCTCCAGACTTAATAATGTATCAACACT 276
DB 261 ATCTCGGGGCGACCAAGCAAGAGAGGGCGCTCTCCAGACTTAATAATGTATCAACACT 202
QY 277 AACCTGTAGAGGGGAGCCCAATCTGAGTCTTCCCGCCTTGGAGACATCGAGCGCGGGA 336
DB 201 AACCTGTAGAGGGGAGCCCAATCTGAGTCTTCCCGCCTTGGAGACATCGAGCGCGGGA 142
QY 337 AGCAGTGGCGGCGGAGGCTGGGCGAGAGAGCTCCAGAGAGGGGCACTGAGCGCTGTGGC 396
DB 141 AGCAGTGGCGGCGGAGGCTGGGCGAGAGAGCTCCAGAGAGGGGCACTGAGCGCTGTGGC 82
QY 397 GCGAGGCTCTGAGACATCGAGAGGACCAAGGAAAGTCTCTCGGGGCGATCTGTAATATA 456
DB 81 GCGAGGCTCTGAGACATCGAGAGGACCAAGGAAAGTCTCTCGGGGCGATCTGTAATATA 22
QY 457 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 477
DB 21 CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 1

```

RESULT 11

CB854121/c
LOCUS

CB854121 573 bp mRNA linear EST 22-APR-2003

DEFINITION UT-CF-DUI-aal-1-16-0-UT s1 UT-CF-DUI Homo sapiens cDNA clone

ACCESSION

CB854121 GI:30044498

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 573)

Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

JOURNAL

MEDLINE

PubMed

COMMENT

Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul.mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA
Sequence: 533-573, >ALU (matched complement)
Seq primer: M13 FORWARD
POLYA=No.

FEATURES

source

Location/Qualifiers
1..573
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UT-CF-DUI-aal-1-16-0-UT"
/tissue_type="Primary Lung Epithelial Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UT-CF-DUI"
/note="Organ: Lung; Vector: pT773-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I.
UT-CF-DUI is a normalized cDNA library containing the following tissue(s): Primary Lung Epithelial Cells The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GGCTGTAGGC.
TAG_SEQ=None found"

ORIGIN

Query Match 75.3%; Score 375; DB 14; Length 573;
Best Local Similarity 100.0%; Pred. No. 9e-146;

Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 108 CTTGAGAGGGCGGCTGTGAGTCTCCAGATATTTGGCCCAAGCGCTTTGGACCTTCT 167
DB 482 CTTGAGAGGGCGGCTGTGAGTCTCCAGATATTTGGCCCAAGCGCTTTGGACCTTCT 423
QY 168 GGACCAAGGCGCCAGAGACTGACGACGAGAGAGAGGGGCTTACCTTTATCTCTGGGGA 227
DB 422 GGACCAAGGCGCCAGAGACTGACGACGAGAGAGAGGGGCTTACCTTTATCTCTGGGGA 363

```

QY 228 CCCACTGCACAAGCAGCGCTCTCCCACTTAATGTATCACTTAACCTGTAGG 287
 DB 362 CCCACTGCACAAGCAGCGCGCTCTCCCACTTAATGTATCACTTAACCTGTAGG 303
 QY 288 GGGAGCCCAATCTGGAATCTCTCCCGCTTGGGACATGACAGCGCGGAAAGCATGCCCC 347
 DB 302 GGGAGCCCAATCTGGAATCTCTCCCGCTTGGGACATGACAGCGCGGAAAGCATGCCCC 243
 QY 348 CCAGGCTGGGCGCAGAGAGCTCCAGAAAGGACATGACGCTGTGGCGAGGCGCTCG 407
 DB 242 CCAGGCTGGGCGCAGAGAGCTCCAGAAAGGACATGACGCTGTGGCGAGGCGCTCG 183
 QY 408 GACATCCGACGACGACGAGAAAGTCTCTGCGGCGATCTGTAATTAACCTTTTCT 467
 DB 182 GACATCCGACGACGACGAGAAAGTCTCTGCGGCGATCTGTAATTAACCTTTTCT 123
 QY 468 TTGTTTAAAAA 482
 DB 122 TTGTTTAAAAA 108

RESULT 12
 A1346155/c 405 bp mRNA linear EST 02-FEB-1999
 LOCUS GP43f12.x1 NCI_CGAP_C08 Homo sapiens cDNA clone IMAGE:1925807 3'
 DEFINITION similar to contains TAR1.t3 TAR1 repetitive element ;, mRNA

ACCESSION A1346155
 VERSION A1346155
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 405)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 Tumor Gene Index
 Tumor Gene Index
 Unpublished (1997)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgapbs-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/BLND at:
 www-bio.lnl.gov/bdrip/image/image.html
 Insert Length: 1488 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 375.
 Location/Qualifiers

FEATURES

source

1..405
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1925807"
 /issue_type="adenocarcinoma"
 /lab_host="DH10B"
 /clone_1lb="NCI_CGAP_C08"
 /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from a colon adenocarcinoma, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match

75.1%; Score 374; DB 9; Length 405;

Best Local Similarity 100.0%; Pred. No. 2.8e-145;
 Matches 374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 73 TGGGTCAAGCTGGGACACCGGGCTGTGTGCTCTCTCGAGAGGGCGGAGTCTC 132
 DB 405 TGGGTCAAGCTGGGACACCGGGCTGTGTGCTCTCTCGAGAGGGCGGAGTCTC 346
 QY 133 CAGTATGTTGGGCCCAAGGCTCTTTCGACCTTTGTGACCAAGCCGCAAGAGCTGACG 192
 DB 345 CAGTATGTTGGGCCCAAGGCTCTTTCGACCTTTGTGACCAAGCCGCAAGAGCTGACG 286
 QY 193 CAGGAGAGAGGGGCTACCTCTTATCTGTGGGACCACTGACAAAGAGGGCGCTTC 252
 DB 285 CAGGAGAGAGGGGCTACCTCTTATCTGTGGGACCACTGACAAAGAGGGCGCTTC 226
 QY 253 CCAGACTTAATGATATCACCCTTAACCTGTGAGGGGACCAATCTGACTCTTCCC 312
 DB 225 CCAGACTTAATGATATCACCCTTAACCTGTGAGGGGACCAATCTGACTCTTCCC 166
 QY 313 GCCTTGGACATCCGAGGCGGGAAGCATGTCGCCGACGCTGGGCGAGAGAGCTCCA 372
 DB 165 GCCTTGGACATCCGAGGCGGGAAGCATGTCGCCGACGCTGGGCGAGAGAGCTCCA 106
 QY 373 GGAAGGCACTGAGGCTGTGTGCGCGAGGCTCGACATCCGACGACACAGGAAAGT 432
 DB 105 GGAAGGCACTGAGGCTGTGTGCGCGAGGCTCGACATCCGACGACACAGGAAAGT 46
 QY 433 CTCTTGGGCGCATC 446
 DB 45 CTCTTGGGCGCATC 32

RESULT 13
 A1660560/c 486 bp mRNA linear EST 18-DEC-1999
 LOCUS we6bD08.x1 Soares Dieckgraebe.colon.NHCD Homo sapiens cDNA clone
 DEFINITION IMAGE:2346231 3' similar to contains TAR1.t3 TAR1 repetitive element ;, mRNA sequence.

ACCESSION A1660560
 VERSION A1660560
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 486)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 Tumor Gene Index
 Tumor Gene Index
 Unpublished (1997)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgapbs-remail.nih.gov
 This clone is available royalty-free through LIND; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 1433 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 447.
 Location/Qualifiers

FEATURES

source

1..486
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2346231"
 /issue_type="colonic mucosa from 3 patients with Crohn's disease"
 /lab_host="DH10B (phage-resistant)"
 /clone_1lb="Soares Dieckgraebe.colon.NHCD"
 /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'] TGTTCACATCTGAAGTGGAGCGCGCGCTCTTTTCTTTTCTTTT 3', double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I

ORIGIN

and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Tissue samples provided by Dr. Brian Dieckgraefe (Washington University, dieck@wustl.edu); colonic mucosa represents a range of disease involvement from moderate to severe Crohn's disease; samples include both perforating (fistulas) and non-perforating samples. Library constructed by Bento Soares and M. Fatima Bonaldo.

Query Match 72.9%; Score 363; DB 9; Length 486;
Best Local Similarity 99.8%; Pred. No. 9, 7e-141;
Matches 413; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 73 TGGGTACGCTGGGACCGGCGTCTTGTGCTCTTCTTGGAGGGCGGTGAGTCTC 132
DB 420 TGGGTACGCTGGGACCGGCGTCTTGTGCTCTTCTTGGAGGGCGGTGAGTCTC 361
QY 133 CAGTATGTTGGCCCGAGGCTCTTGTGACCTTGTGACCAAGCGCCAGAGCTGACG 192
DB 360 CAGTATGTTGGCCCGAGGCTCTTGTGACCTTGTGACCAAGCGCCAGAGCTGACG 301
QY 193 CAGAGAGAGGGGGCTCACTCTTATCTCTGCGACCCCACTGACAGAGAGCGCTCTC 252
DB 300 CAGAGAGAGGGGGCTCACTCTTATCTCTGCGACCCCACTGACAGAGAGCGCTCTC 241
QY 253 CCAACTTAAATGATATACACTTACCTGTGAGGGGAGCCCAATCTGACTCTTCCCC 312
DB 240 CCAACTTAAATGATATACACTTACCTGTGAGGGGAGCCCAATCTGACTCTTCCCC 181
QY 313 GCTTGGAGACATCGAGGCGCGGAGAGAGTGCCTGCGAGGCTGAGGCGAGAGAGCTCA 372
DB 180 GCTTGGAGACATCGAGGCGCGGAGAGAGTGCCTGCGAGGCTGAGGCGAGAGAGCTCA 121
QY 373 GGAAGGGCACTGAGGCGCTGTGCGCGGAGGCTCTGGAATCCGAGGACAGGAAAGT 432
DB 120 GGAAGGGCACTGAGGCGCTGTGCGCGGAGGCTCTGGAATCCGAGGACAGGAAAGT 61
QY 433 CTCCTGGGGCGGCTGTGTAATTAACCTTTTCTTTTCTTTTAAATTAATTAATTA 486
DB 60 CTCCTGGGGCGGCTGTGTAATTAACCTTTTCTTTTCTTTTAAATTAATTAATTA 7

RESULT 14
LOCUS CB305399 737 bp mRNA linear EST 04-MAR-2003
DEFINITION UI-CF-EN1-aed-m-17-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone
ACCESSION CB305399
VERSION CB305399.1 GI:28845910
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 737)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: McCray, PB
McCrays Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@iowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

FEATURES
source

Clone Distribution: Researchers may obtain clones from Research Genetics (www.reagen.com) or from Open Biosystems (www.openbiosystems.com).
The following repetitive elements were found in this cDNA sequence: 29-138, >MIR#SINE/MIR (matched complement) 195-290, >ALU 223-330, >ALU
Seq primer: M13 FORWARD
POLYA=yes.

ORIGIN

Query Match 69.9%; Score 348; DB 14; Length 737;
Best Local Similarity 100.0%; Pred. No. 1, 4e-134;
Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 GGCCTCTGCTGCTCTTCTCTGAGAGGGCCGCTGAGTCTCCAGTATGTTGGCCACG 150
DB 353 GGCCTCTGCTGCTCTTCTCTGAGAGGGCCGCTGAGTCTCCAGTATGTTGGCCACG 412
QY 151 GCTTTGGACACCTTCTTGAACCAAGGCCCAAGACTGACCCAGAGAGAGGGGCTCA 210
DB 413 GCTTTGGACACCTTCTTGAACCAAGGCCCAAGACTGACCCAGAGAGAGGGGCTCA 472
QY 211 CCTCTTATCTGCGGACCCAGTGCACAGAGGAGGCGCTCTCCAGACTTAAATATATC 270
DB 473 CCTCTTATCTGCGGACCCAGTGCACAGAGGAGGCGCTCTCCAGACTTAAATATATC 532
QY 271 ACCACTAATCTGTGAGGGGAGCCCAATCTGAGTCTCTTCCCGCTTGGAGCAATGCGAG 330
DB 533 ACCACTAATCTGTGAGGGGAGCCCAATCTGAGTCTCTTCCCGCTTGGAGCAATGCGAG 592
QY 331 CCGGAGAGAGTGGCGCGGCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 390
DB 593 CCGGAGAGAGTGGCGCGGCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 652
QY 391 GCTGGCGAGAGGCTTGGACATCGAGAGGACAGGAGGAGGAGGAGGAGGAGGAGGAGG 438
DB 653 GCTGGCGAGAGGCTTGGACATCGAGAGGACAGGAGGAGGAGGAGGAGGAGGAGGAGG 700

RESULT 15
LOCUS AA573825/c 466 bp mRNA linear EST 12-SEP-1997
DEFINITION AA573825 nk80a11.s1 NCT_CGAB_Co2 Homo sapiens cDNA clone IMAGE:1012892 3',
mRNA sequence.

ACCESSION AA573825
 VERSION AA573825.1 GI:2348340
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 466)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgapbs-remail.nih.gov
 Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Stratagene, Inc., David B. Krizman,
 Ph.D.

CDNA Library Arraying: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
www-bio.lnl.gov/bbrp/image/image.html
 Insert length: 741 Std Error: 0.00
 Seq primer: -40m3 fwd. ET from Amersham
 High quality sequence stop: 456.
 Location/Qualifiers
 1..466

/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1012892"
 /tissue_type="tumor"
 /lab_host="GOLR (kanamycin resistant)"
 /clone_1ib="NCI-CGAP_CO2"
 /note="Organ: colon; Vector: Bluescript SK-; Site: 1;
 EcoRI; Site: 2; XhoI; Cloned unidirectionally. Primer:
 Oligo dt. Bulk colon villous adenoma. 5' adaptor sequence:
 5' GAATTCGACGAG 3' 3' adaptor sequence: 5'
 CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

ORIGIN

Query Match 66.7%; Score 332; DB 9; Length 466;
 Best Local Similarity 100.0%; Pred. No. 8.3e-128;
 Matches 332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 60 GAGCGCGGCTTCTGGGTACGCTGCAACCGGCGTCTGCTTCTTCTCGAGGGGC 119
 DB 417 GAGCGCGGCTTCTGGGTACGCTGCAACCGGCGTCTGCTTCTTCTCGAGGGGC 358
 QY 120 CGTGTGAGTCTCCAGTATGTTGGGCCCAAGCGCTTCTTGCACCTTCTGACCAAGCGC 179
 DB 357 CGTGTGAGTCTCCAGTATGTTGGGCCCAAGCGCTTCTTGCACCTTCTGACCAAGCGC 298
 QY 180 CAAGACTGAGCGAGGAGAGAGGGGGCTCACTCTTACTCTCGGCGGACCACTGCACAA 239
 DB 297 CAAGACTGAGCGAGGAGAGGGGGCTCACTCTTACTCTCGGCGGACCACTGCACAA 238
 QY 240 GCAGCGCGCTTCTCCGACTTAAATGTATCACTAACCCTGTGAGGGGAGCCCAATCT 299
 DB 237 GCAGCGCGCTTCTCCGACTTAAATGTATCACTAACCCTGTGAGGGGAGCCCAATCT 178
 QY 300 GGACTCCTTCCCGCTTGGGACATGCAAGGCCGGGAAAGCAATGCCCGCCAGGCTTGGGC 359
 DB 177 GGACTCCTTCCCGCTTGGGACATGCAAGGCCGGGAAAGCAATGCCCGCCAGGCTTGGGC 118
 QY 360 CAGAGAGCTCCAGAGAGGCACTGAGCGCTG 391
 DB 117 CAGAGAGCTCCAGAGAGGCACTGAGCGCTG 86

Search completed: February 24, 2004, 03:21:11
 Job time: 1268.96 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 19:34:53 ; Search time 1965.03 Seconds

(without alignments)
10984.465 Million cell updates/sec

Title: US-09-936-456-3

Perfect score: 1 ctcctagcgcgcgcctctgcc.....aaaaaaaaataaagtcgacc 498

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenBank: 1: gb ba: 2: gb htg: 3: gb in: 4: gb cm: 5: gb ov: 6: gb pat: 7: gb ph: 8: gb pl: 9: gb pr: 10: gb ro: 11: gb sts: 12: gb sy: 13: gb un: 14: gb vi: 15: em ba: 16: em fun: 17: em hum: 18: em in: 19: em mu: 20: em om: 21: em or: 22: em ov: 23: em pac: 24: em ph: 25: em pl: 26: em ro: 27: em sts: 28: em un: 29: em vi: 30: em htg hum: 31: em htg inv: 32: em htg other: 33: em htg mus: 34: em htg pln: 35: em htg rod: 36: em htg mam: 37: em htg vrt: 38: em sy: 39: em htgo hum: 40: em htgo mus: 41: em htgo other:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Query Match	Length	DB	ID	Description
1	498	100.0	498	6	AX035348	AX035348 Sequence
2	465.4	93.5	1441	6	AX035346	AX035346 Sequence
3	463.8	93.1	1740	6	HSX806241	BX537581 Homo sapi
4	398.4	80.0	2684	6	AX714511	AX714511 Sequence
5	398.4	80.0	2684	2	AX056896	AX056896 Homo sapi
6	398.4	80.0	156534	2	AC009700	AC009700 Homo sapi
7	398.4	80.0	166937	2	AC012255	AC012255 Homo sapi
8	398.4	80.0	171444	2	AC087790	AC087790 Homo sapi
9	398.4	80.0	181312	9	AC091117	AC091117 Homo sapi
10	217	43.6	217	9	HS183412F	Z57552 H. sapiens C
11	199.8	40.1	218	9	HS183412R	Z59554 H. sapiens C
12	166	33.3	1923	9	BC029819	BC029819 Homo sapi
13	160.6	32.2	1354	10	BC031111	BC031111 Mus muscu
14	117.8	23.7	229583	10	AL844566	AL844566 Mouse DNA
15	113.6	22.8	5851	4	AF547266	AF547266 Sus scrof
16	95	19.1	231868	2	AC118124	AC118124 Rattus no
17	50.2	10.1	125020	9	AF429315	AF429315 Homo sapi
18	48.2	9.7	125020	9	AF429315	AF429315 Homo sapi
19	46	9.2	1762	9	BC064138	BC064138 Homo sapi
20	46	9.2	3180	6	AB882116	AB882116 Sequence
21	46	9.2	3180	6	BD159535	BD159535 Primer fo
22	46	9.2	3180	9	AK022395	AK022395 Homo sapi
23	46	9.2	119347	9	AC020934	AC020934 Homo sapi
24	46	9.2	147750	9	AC018761	AC018761 Homo sapi
25	41.6	8.4	299425	1	AP005049	AP005049 streptomy
26	40.2	8.1	33676	1	SCARDIGN	X84374 Saccharochr
27	40	8.0	3728	1	AB110077	AB110077 Streptomy
28	39.8	8.0	52359	2	AC010772	AC010772 Homo sapi
29	39.8	8.0	244105	2	AC098897	AC098897 Rattus no
30	39.6	8.0	160513	2	AC113563	AC113563 Canis fam
31	39.2	7.9	42338	8	HDMO16F8	D85922 Homo sapien
32	38	7.8	117974	8	AC105730	AC105730 Oryza sat
33	38	7.8	166282	8	AC134233	AC134233 Oryza sat
34	38	7.8	182010	2	AC016767	AC016767 Homo sapi
35	38.8	7.8	2499	6	BD180276	BD180276 Highly th
36	38.8	7.8	34721	9	HS315055	AL031708 Human DNA
37	38.8	7.8	270150	9	AE006639	AE006639 Homo sapi
38	38.6	7.8	2000	6	AX655393	AX655393 Sequence
39	38.6	7.8	166889	2	AC129938	AC129938 Mus muscu
40	38.6	7.8	218588	10	AC113497	AC113497 Mus muscu
41	38.6	7.8	299300	1	AP005026	AP005026 Streptomy
42	38.2	7.7	1094	4	SSC440724	AF440724 Sus scrof
43	38.2	7.7	187080	9	AP000977	AP000977 Homo sapi
44	38.2	7.7	187917	2	AC025321	AC025321 Homo sapi
45	38.2	7.7	201509	2	AC145356	AC145356 Gorilla g

ALIGNMENTS

RESULT 1
LOCUS AX035348 498 bp DNA linear PAT 15-NOV-2000
DEFINITION Sequence 3 from Patent WO0053748.
ACCESSION AX035348
VERSION AX035348.1 GI:11191065
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1
AUTHORS Bruck,C.E., Vinals,Y.D., Coche,T. and Cassart,J.P.
TITLE Novel compounds
JOURNAL Patent: WO 0053748-A 3 14-SEP-2000;

BRUCK CLAUDE ELVIRE MARIE (BE) ; SMITHKLINE BEECHAM BIOLOG (BE) ;
VINALS Y DE BASOLS CARLOTA (BE) ; COCHE THIERRY (BE) ; CASSART
JEAN POL (BE)

FEATURES
source
Location/Qualifiers
1..498
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 498; DB 6; Length 498;

Best Local Similarity 100.0%; Pred. No. 9,6e-113; Mismatches 0; Indels 0; Gaps 0;

Matches 488; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCTAGCGTGGCGCTCTGCGCGCTCCGCTTACAGCTCTCCGCGCTACACCTAGTAGG 60
DB 1 CTCTAGCGTGGCGCTCTGCGCGCTCCGCTTACAGCTCTCCGCGCTACACCTAGTAGG 60
QY 61 AGGCGCGCTCTGCGCGCTCTGCGCGCTCCGCTTACAGCTCTCCGCGCTACACCTAGTAGG 120
DB 61 AGGCGCGCTCTGCGCGCTCTGCGCGCTCCGCTTACAGCTCTCCGCGCTACACCTAGTAGG 120
QY 121 GTGTGAGTCTCCAGTATGTTGCGCGCTCCGCTTACAGCTCTCCGCGCTACACCTAGTAGG 180
DB 121 GTGTGAGTCTCCAGTATGTTGCGCGCTCCGCTTACAGCTCTCCGCGCTACACCTAGTAGG 180
QY 181 AAGGACTGACGACGAG 240
DB 181 AAGGACTGACGACGAG 240
QY 241 CAGGCGCTCTCCGCGCTCTGCGCGCTCCGCTTACAGCTCTCCGCGCTACACCTAGTAGG 300
DB 241 CAGGCGCTCTCCGCGCTCTGCGCGCTCCGCTTACAGCTCTCCGCGCTACACCTAGTAGG 300
QY 301 GACTCTCTCTCCGCGCTCTGCGCGCTCCGCTTACAGCTCTCCGCGCTACACCTAGTAGG 360
DB 301 GACTCTCTCTCCGCGCTCTGCGCGCTCCGCTTACAGCTCTCCGCGCTACACCTAGTAGG 360
QY 361 AGGAGAGCTCCAG 420
DB 361 AGGAGAGCTCCAG 420
QY 421 ACCAGGAGAAAGTCTCTGCGCGCTCTGCGCGCTCCGCTTACAGCTCTCCGCGCTACACCTAGTAGG 480
DB 421 ACCAGGAGAAAGTCTCTGCGCGCTCTGCGCGCTCCGCTTACAGCTCTCCGCGCTACACCTAGTAGG 480
QY 481 AAAAAATAAAGTCGAGC 498
DB 481 AAAAAATAAAGTCGAGC 498

RESULT 2
AX035346 1441 bp DNA linear PAT 15-NOV-2000
LOCUS Sequence 1 from Patent WO0053748.
DEFINITION AX035346
ACCESSION AX035346
VERSION AX035346.1 GI:11191064
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
Bruck, C.E., Vinals, Y.D., Coche, T. and Cassart, J.P.
Novel compounds
Patent: WO 0053748-A 1 14-SEP-2000;
SMITHKLINE BEECHAM BIOLOG (BE) ;
BRUCK CLAUDE ELVIRE MARIE (BE) ;
VINALS Y DE BASOLS CARLOTA (BE) ; COCHE THIERRY (BE) ; CASSART
JEAN POL (BE)

FEATURES
source
1..1441
/organism="Homo sapiens"
/mol_type="unassigned DNA"

/db_xref="taxon:9606"

Query Match 93.5%; Score 465.4; DB 6; Length 1441;
Best Local Similarity 99.4%; Pred. No. 1.2e-104;
Matches 488; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 1 CTCTAGCGTGGCGCTCTGCGCGCTCCGCTTACAGCTCTCCGCGCTACACCTAGTAGG 60
DB 939 CTCTAGCGTGGCGCTCTGCGCGCTCCGCTTACAGCTCTCCGCGCTACACCTAGTAGG 997
QY 61 AGGCGCGCTCTGCGCGCTCTGCGCGCTCCGCTTACAGCTCTCCGCGCTACACCTAGTAGG 120
DB 998 -GGCGCGCTCTGCGCGCTCTGCGCGCTCCGCTTACAGCTCTCCGCGCTACACCTAGTAGG 1056
QY 121 GTGTGAGTCTCCAGTATGTTGCGCGCTCCGCTTACAGCTCTCCGCGCTACACCTAGTAGG 180
DB 1057 GTGTGAGTCTCCAGTATGTTGCGCGCTCCGCTTACAGCTCTCCGCGCTACACCTAGTAGG 1116
QY 181 AAGGACTGACGACGAG 240
DB 1117 AAGGACTGACGACGAG 1176
QY 241 CAGGCGCTCTCCGCGCTCTGCGCGCTCCGCTTACAGCTCTCCGCGCTACACCTAGTAGG 300
DB 1177 CAGGCGCTCTCCGCGCTCTGCGCGCTCCGCTTACAGCTCTCCGCGCTACACCTAGTAGG 1236
QY 301 GACTCTCTCTCCGCGCTCTGCGCGCTCCGCTTACAGCTCTCCGCGCTACACCTAGTAGG 360
DB 1237 GACTCTCTCTCCGCGCTCTGCGCGCTCCGCTTACAGCTCTCCGCGCTACACCTAGTAGG 1296
QY 361 AGGAGAGCTCCAG 420
DB 1297 AGGAGAGCTCCAG 1356
QY 421 ACCAGGAGAAAGTCTCTGCGCGCTCTGCGCGCTCCGCTTACAGCTCTCCGCGCTACACCTAGTAGG 480
DB 1357 ACCAGGAGAAAGTCTCTGCGCGCTCTGCGCGCTCCGCTTACAGCTCTCCGCGCTACACCTAGTAGG 1416
QY 481 AAAAAATAAAGTCGAGC 491
DB 1417 AAAAAATAAAGTCGAGC 1427

RESULT 3
HSMB06241 1740 bp mRNA linear PRI 17-JUN-2003
LOCUS Homo sapiens mRNA, cDNA DKFZp686C04213 (from clone DKFZp686C04213).
DEFINITION BX537581
ACCESSION BX537581
VERSION BX537581.1 GI:31873659
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
Bloeker, H., Boecker, M., Mewes, H.W., Weil, B., Amid, C., Osanger, A.,
Fodor, G., Han, M. and Wiemann, S.
Direct Submission
Submitted (17-JUN-2003) MTPS, Ingolstaedter Landstr.1, D-85764
Neuerberg, GERMANY
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
sequenced by GBF (National Research Centre for Biotechnology Ltd.,
Braunschweig/Germany) within the cDNA sequencing consortium of the
German Genome Project.
This clone (DKFZp686C04213) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at <http://mips.gsf.de/proj/cDNA/>.

FEATURES
source
1..1740
Location/Qualifiers


```

/organism="Homo sapiens"
/mol_type="rRNA"
/db_xref="taxon:9606"
/map="15q15"
/cloname="DKFZp686C04213"
/tissue_type="human rectum tumor"
/clonelib="686 (synonym: h1cc3) . Vector pSport1_sf; host
DH10B; sites SfilA + SfilB"
/dev_stage="adult"
1694 . 1699
polya_signal
polya_site
1720
ORIGIN

```

Query Match	93.1%;	Score 463.8;	DB 9;	Length 1740;
Best Local Similarity	99.2%;	Pred. No. 3e-104;		
Matches 487;	Conservative	0;	Mismatches 2;	Indels 2;
				Gaps 2;

QY	CTTAGAGTGC	CGCCTCTG	CCCGCCTCC	CCGCTAGG	CTCTCG	CGCTAC	CACTCA	CGTACG	30
Db	CTTAGAGTGC	CGCCTCTG	CCCGCCTCC	CCGCTAGG	CTCTCG	CGCTAC	CACTCA	CGTACG	13046
QY	AGCGCGCCTT	CTG	GGTCA	CGCTGG	CAACCG	CGCTCT	GT	GTGCTCTT	120
Db	AGCGCGCCTT	CTG	GGTCA	CGCTGG	CAACCG	CGCTCT	GT	GTGCTCTT	1305
QY	GTGATGATCT	CAGATAT	TTT	CG	CGCCAG	CGCTCT	GT	GGACACCTT	180
Db	GTGATGATCT	CAGATAT	TTT	CG	CGCCAG	CGCTCT	GT	GGACACCTT	1364
QY	GTGATGATCT	CAGATAT	TTT	CG	CGCCAG	CGCTCT	GT	GGACACCTT	1422
Db	GTGATGATCT	CAGATAT	TTT	CG	CGCCAG	CGCTCT	GT	GGACACCTT	1422
QY	AAGGACTG	CAAGCA	AGGAGAG	AGGGGG	CTC	ACTT	T	TCTCGG	240
Db	AAGGACTG	CAAGCA	AGGAGAG	AGGGGG	CTC	ACTT	T	TCTCGG	1483
QY	CAGAGCGCTCT	CCCA	CACTTAA	ATAT	ATAT	ATCA	CACTA	ACCTGT	300
Db	CAGAGCGCTCT	CCCA	CACTTAA	ATAT	ATAT	ATCA	CACTA	ACCTGT	1544
QY	GACTCTT	CCCGCCTT	GG	CAATG	CAAG	CGCGG	AAACAT	GTG	360
Db	GACTCTT	CCCGCCTT	GG	CAATG	CAAG	CGCGG	AAACAT	GTG	1544
QY	AGGAGAGCTC	ACAGAA	GGGCA	CTG	AG	CGCTCT	GG	CGCGCT	420
Db	AGGAGAGCTC	ACAGAA	GGGCA	CTG	AG	CGCTCT	GG	CGCGCT	1604
QY	ACCAGGAAAG	TCTCT	CGG	CGCAT	CT	TAAAT	TAA	ACTTTT	480
Db	ACCAGGAAAG	TCTCT	CGG	CGCAT	CT	TAAAT	TAA	ACTTTT	1722
QY	AAAAAT	TAAAA	491						
Db	AAAAAT	TAAAA	1734						

RESULT	4
AX714511/c	
LOCUS	AX714511
DEFINITION	Sequence 1195 from Patent EP1293569.
ACCESSION	AX714511
VERSION	AX714511.1
KEYWORDS	GI:29889464
SOURCE	.
ORGANISM	Homo sapiens (human)
	Homo sapiens
	PAT 15-APR-2003
	DNA linear

REFERENCE AUTHORS

Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, Y., Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R., Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and Masuhbo, Y.

TITLE Full-length cDNAs
JOURNAL Patent: EP 1293569-A 1195 19-MAR-2003;
Helix Research Institute (JP) ; Research Association for

```

FEATURES
  source
    Biotechnology (JF)
    Location/Qualifiers
      1..2684
        /organism="Homo sapiens"
        /mol_type="unassigned DNA"
        /db_xref="taxon:9606"
ORIGIN
```

Query Match	80.0%;	Score 398.4;	DB 6;	Length 2684;
Best Local Similarity	95.1%;	Pred. No. 5e-88;		
Matches 411; Conservative	0;	Mismatches 21;	Indels 0;	Gaps 0;

Oy	5	ACTCAGTATGAGACGCGGCTTCTGAGTCAAGCTGGCAACCGAGGTCCTGCTGCTCTTCT	110
Db	2383	ACTCCACATGCGCTCTCTTCTTTGGATCCCAACCGCACAGGGGCTCTGTGGCTCTTCT	232
Oy	111	CGAGAGGGCCGTGTGAATCTCCAGTATGTTCCGACCAGCTCTTTCGACCCCTTCTGA	170
Db	2323	CGAGAGGGCCGTGTGAATCTCCAGTATGTTCCGACCAGGCTCTTTCGACCCCTTCTGA	226
Oy	171	CCAAAGCGCGACAGACTGACGACGAGAGAGAGGGGGCTACCTTATCTCGGCGAGCC	230
Db	2263	CCAAAGCGCGACAGACTGACGACGAGAGAGGGGGCTACCTTATCTCGGCGAGCC	220
Oy	231	ACTGCAAGACGAGCGGCTCTCCGACTTAAATGTATACCACTAACCTGTGAGGGG	290
Db	2203	ACTGCAAGACGAGCGGCTCTCCGACTTAAATGTATACCACTAACCTGTGAGGGG	214
Oy	291	ACCCATCTGGACTCTTCCCGGCTTTGGGACATCGAGGCGGGAGAGAGTCCCGCA	350
Db	2143	ACCCATCTGGACTCTTCCCGGCTTTGGGACATCGAGGCGGGAGAGAGTCCCGCA	208
Oy	351	GGGCTGGGCGCAGAGAGCTCCAGAAAGGCACTGAGGCTGTGCGCGCAGGCGCTCGAC	410
Db	2083	GGGCTGGGCGCAGAGAGCTCCAGAAAGGCACTGAGGCTGTGCGCGCAGGCGCTCGAC	202
Oy	411	ATCCGAGGACCCAGGAAAGTCTCTCGGGCGATCTGTAAATPAACTTTTTTTCTTT	470
Db	2023	ATCCGAGGACCCAGGAAAGTCTCTCGGGCGATCTGTAAATPAACTTTTTTTCTTT	196
Oy	471	GTTTTAAAAA 482	
Db	1963	GTTTTAAAAA 1952	

RESULT	5
LOCUS	AK056896/c
DEFINITION	Homo sapiens cDNA FLJ33334 f.1.s, clone PROST2005426.
ACCESSION	AK056896
VERSION	AK056896.1 GI:16552419
KEYWORDS	oligo caping; f.1.s (full insert sequence).
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE AUTHORS

AUTHORS
 Imahashi, T., Kanehori, K., Yosida, M., Matanabe, S., Ishida, S.,
 Ono, Y., Houta, T., Hirooka, S., Murakawa, K., Takiguchi, S.,
 Kusano, Y., Matanabe, M., Fujimori, K., Tanai, H., Ishida, M.,
 Yamashita, H., Chiba, Y., Sugiyama, T., Tate, R., Otsuki, T., Sato, H.,
 Makatsugu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai, Hio, Y.,
 Saito, K., Nishikawa, T., Kimura, K., Matsuo, K., Takahashi, Y.,
 Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Takamura-Fujii, A.,
 Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S.,
 Nagahara, K., Masuko, Y., Nagai, K. and Isegai, T.

TITLE
 NEDO human cDNA sequencing project

JOURNAL
 Unpublished

REFERENCE
 2 (bases 1 to 2684)

NUMBERS
 Accelink T
 Accelink T
 Accelink T

TITLE Direct Submission
JOURNAL Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1552-3 Yana, Kisarazu, Chiba 292-0812, Japan

COMMENT

(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
 NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center; National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

FEATURES

source

Location/Qualifiers
 1. 2684
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="PROST2005426"
 /issue_type="prostate"
 /clone_lib="PROST2"
 /note="cloning vector: pME18SFL3"
 56..952
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAB71304.1"
 /db_xref="GI:16552420"
 /translation="MATLGHFFPYAGSKPTFPMQDTLTIIMFTLTAATFVILPG
 IGRKRLFWLRLVNTSLFGAIIIGTPVQOINTEINNEFTMLGNVAEYKALE
 KGLPDPVYLAEKFTPRSPCGLYRQYRLAGHYTAMNVMVAFICLLANVMSPVLY
 GGYMLATGIFOLLALFFSMATSLTSPCHLASVLAHTHGAPAFITLTGLCTL
 LGLAAVAARMOPHRLKAFPNQSVEDPMLMSEEGLLSPRRMSADSPKQDPL
 SRASSTKAYCKRAHPKDPCL"

CDS

51 ACTCATAGAGAGCGCCCTTCTGGGTACACGCTGGCAACGCGCTGTCCTTCTTCT
 2383 ACTCCACATGCCCCCTTTCTTTCATCCACCGCCACAGCGCTGTCCTTCTTCT
 111 CGAGAGGCGCGTGTAGTCTCCAGTATGTTGCGCCACGCTCTTCCGACCTTCTGGA
 2323 CGAGAGGCGCGTGTAGTCTCCAGTATGTTGCGCCACGCTCTTCCGACCTTCTGGA
 171 CCAAGCGCCAGAGATCGACCGAGAGAGAGGCGCTCACTCTTATCTCGGCAACC
 2263 CCAAGCGCCAGAGATCGACCGAGAGAGAGGCGCTCACTCTTATCTCGGCAACC
 2231 ACTGACACAGAGCGCGCTCTCCAGACTTAATGTATACCACTAAGCTGTAGAGGGG
 2203 ACTGACACAGAGCGCGCTCTCCAGACTTAATGTATACCACTAAGCTGTAGAGGGG
 291 ACCCAATCTGAGCTCTTCTCCGCGCTTGGGACATGACAGCGCGGAGACAGTGCCTGCA
 2143 ACCCAATCTGAGCTCTTCTCCGCGCTTGGGACATGACAGCGCGGAGACAGTGCCTGCA
 351 GGCCTGGGAGAGAGCTCCAGAAAGGACATGACAGCGCTGCGAGAGAGCTCGGAC
 2083 GGCCTGGGAGAGAGCTCCAGAAAGGACATGACAGCGCTGCGAGAGAGCTCGGAC
 411 ATCCGAGGACCAAGGAAAGTCTCTCGGAGCATGTGTAATAAATCTTTTCTTT
 2023 ATCCGAGGACCAAGGAAAGTCTCTCGGAGCATGTGTAATAAATCTTTTCTTT
 471 GTTTTAAAAA 482
 1963 GTTTTAAAAA 1952

ORIGIN

Query Match 80.0%; Score 388.4; DB 9; Length 2684;
 Best Local Similarity 95.1%; Pred. No. 5e-88;
 Matches 411; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

RESULT 6
 AC009700
 LOCUS
 DEFINITION

AC009700 156534 bp DNA linear HTG 20-APR-2000
 Homo sapiens chromosome 15 clone RP11-163P10 map 15, WORKING DRAFT
 SEQUENCE, 15 unordered pieces.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS

TITLE
 JOURNAL

COMMENT

AC009700
 AC009700.4 GI:7622346
 HTG; HTGS_PHASE1; HTGS_DRAFT.
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 156534)
 Birren, B., Linton, L., Nusbaum, C., Lander, B., Allen, N., Anderson, M.,
 Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A.,
 Cagle, A., Cerny, J., Colangelo, M., Collins, S., Colymore, A.,
 Cooke, P., DeRubeis, K., Depireux, E., Devon, K., Dewar, K.,
 Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C.,
 Funke, R., Gage, D., Galagan, J., Gardy, S., Gilbert, D., Grant, G.,
 Hagos, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kann, L.,
 Karatas, A., Lehoczy, J., Liu, C., Locke, K., Macdonald, P.,
 Margulis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J.,
 Meidlin, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,
 Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,
 Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,
 Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,
 Testaye, S., Tortorella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,
 Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.
 Direct Submission
 Submitted (28-AUG-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 20, 2000 this sequence version replaced gi:6056272.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center -----
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIRB
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information -----
 Center project name: 163.P.10
 Center clone name: 163.P.10
 ----- Summary Statistics -----
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 108837 bases at least Q40
 Consensus quality: 110230 bases at least Q30
 Consensus quality: 145227 bases at least Q20
 Insert size: 15700; agarose-fp
 Insert size: 155134; sum-of-ctnigs
 Quality coverage: 3.2 in Q20 bases; sum-of-ctnigs
 Quality coverage: 3.3 in Q20 bases; sum-of-ctnigs
 ----- NOTE: This is a 'working draft' sequence. It currently
 consists of 15 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.
 1 1169: contig of 1169 bp in length
 * 1170 1269: gap of 100 bp
 * 1270 3335: contig of 2066 bp in length
 * 3336 3435: gap of 100 bp
 * 3436 5698: contig of 2263 bp in length
 * 5699 5798: gap of 100 bp
 * 5799 9216: contig of 3418 bp in length
 * 9217 9316: gap of 100 bp
 * 9317 15633: contig of 6317 bp in length
 * 15634 15733: gap of 100 bp
 * 15734 21746: contig of 6013 bp in length

[illegible]

```

source
1. .156534
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-163P10"
1. .1169
/clone.lib="RPC1-11 Human Male BAC"
/misc_feature
1270. .335
/note="assembly_fragment"
/misc_feature
3436. .5698
/note="assembly_fragment"
/misc_feature
5799. .9216
/note="assembly_fragment"
/misc_feature
9317. .1563
/note="assembly_fragment"
/misc_feature
15734. .21746
/note="assembly_fragment"
/misc_feature
21847. .30351
/note="assembly_fragment"
/misc_feature
30452. .36509
/note="assembly_fragment"
/misc_feature
36610. .45279
/note="assembly_fragment"
/misc_feature
45380. .56851
/note="assembly_fragment"
/misc_feature
56952. .67374
/note="assembly_fragment"
/misc_feature
67475. .79534
/note="assembly_fragment"
/misc_feature
79635. .95432
/note="assembly_fragment"
/misc_feature
95533. .120801
/note="assembly_fragment"
/misc_feature
120902. .156534
/note="assembly_fragment"

```

Query Match	80.0%	Score 398.4	DB 2	Length 156534
Best Local Similarity	95.1%	Pred. No. 5	Se-88	
Matches	411	Conservative	0	Mismatches 21
				Indels 0
				Gaps 0
QY	51	ACTCAGTACGAGCGCCGCTTTCGGTTCAGCTGGCAACGGGGTCTCTGCTCTTCT	110	
Db	96091	ACTCCACATGCGCCCTCTTCTTTCGATGCCACCGCACAGGGGTCTGTGCTCTTCT	96150	
QY	111	CGAGAGGGCCGTGTGATGATCTCCAGTATGTTGGCCAGCGCTCTTTCGACCCCTTGTGA	170	
Db	96151	CGAGAGGGCCGTGTGATGATCTCCAGTATGTTGGCCAGCGCTCTTTCGACCCCTTGTGA	96210	

QY	17	CCAAAGGCCAAGACATCGAGCAGGAGAGAGGGGGCTCACTCTTATCTTCGGGACCC	230
Db	96211	CCAAAGGCCAAGACATCGAGCAGGAGAGAGGGGGCTCACTCTTATCTTCGGGACCC	96270
QY	231	ACTGCACACAGCAGCGCGCTCTCCAGACTTTAAATGTATCACCACTAACCTGTAGGGG	290
Db	96271	ACTGCACACAGCAGCGCGCTCTCCAGACTTTAAATGTATCACCACTAACCTGTAGGGG	96330
QY	291	ACCCATCTGACTCTCCTCCCGCCTTGGGACATGCACAGCGGGAGCAGTGGCCCA	350
Db	96331	ACCCATCTGACTCTCCTCCCGCCTTGGGACATGCACAGCGGGAGCAGTGGCCCA	96390
QY	351	GGCTTGGGCCAGGAGAGCTTCAAGAAAGGCACTGACGCTGCTGGCGCAGGCTTCGAC	410
Db	96391	GGCTTGGGCCAGGAGAGCTTCAAGAAAGGCACTGACGCTGCTGGCGCAGGCTTCGAC	96450
QY	411	ATCGCAGGACACAGGAAAGTCTCCTGGGGCGATCTGTAAATAAACCTTTTTCTTT	470
Db	96451	ATCGCAGGACACAGGAAAGTCTCCTGGGGCGATCTGTAAATAAACCTTTTTCTTT	96510
QY	471	GTTTTAAAAA 482	
Db	96511	GTTTTAAAAA 96522	

[illegible]

Center Whitehead Institute/ MIT Center for Genome Research
Center code: M13
Web site: <http://www.ssg.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: L3737
Center clone name: 109_D_20
Summary Statistics
Sequencing vector: M13, M77815; 100% of reads

Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 15380 bases at least Q40
Consensus quality: 16066 bases at least Q30
Consensus quality: 163275 bases at least Q20
Insert size: 164837; sum-of-contigs
Quality coverage: 4.0 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
consists of 22 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 2087: contig of 2087 bp in length
2088 2187: gap of 100 bp
2188 3626: contig of 1439 bp in length
3627 3726: gap of 100 bp
3727 4845: contig of 1119 bp in length
4846 4945: gap of 100 bp
4946 6702: contig of 1757 bp in length
6703 6802: gap of 100 bp
6803 8326: contig of 1524 bp in length
8327 8426: gap of 100 bp
8427 10484: contig of 2068 bp in length
10485 10584: gap of 100 bp
10595 13215: contig of 2621 bp in length
13216 13315: gap of 100 bp
13316 16451: contig of 3136 bp in length
16452 16551: gap of 100 bp
16552 20843: contig of 4292 bp in length
20844 20943: gap of 100 bp
20944 24940: contig of 3897 bp in length
24941 25041: gap of 100 bp
25041 29901: contig of 4861 bp in length
29902 30001: gap of 100 bp
30002 34189: contig of 4188 bp in length
34190 34289: gap of 100 bp
34290 38386: contig of 4097 bp in length
38387 38486: gap of 100 bp
38487 45827: contig of 7341 bp in length
45828 45928: gap of 100 bp
45929 51729: contig of 5802 bp in length
51730 51829: gap of 100 bp
51830 60123: contig of 8294 bp in length
60124 60223: gap of 100 bp
60224 73532: contig of 13309 bp in length
73533 73632: gap of 100 bp
73633 87433: contig of 13801 bp in length
87434 87533: gap of 100 bp
87534 101708: contig of 14175 bp in length
101709 101808: gap of 100 bp
101809 121193: contig of 19385 bp in length
121194 121293: gap of 100 bp
121294 139051: contig of 17758 bp in length
139052 139151: gap of 100 bp
139152 166937: contig of 27786 bp in length.

FEATURES

source
1. 166937
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-109D20"
/clone_lib="RP11 Human Male BAC"
1. 2087
/note="assembly_fragment"
misc_feature
2188..3626
/note="assembly_fragment"
misc_feature
3727..4845

ORIGIN

Query Match 80.0%; Score 398.4; DB 2; Length 166937;
Best Local Similarity 95.1%; Pred No. 5.5e-88;
Matches 411; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

51 ACTGAGTACGAGCGCCGCTTCTGGGTACGCTGGCAACCGGCTCTGTCTTCTCT 110
161043 ACTCCACATGCGCTCTCTTCTTTCGATCCCAACCGCACAGGCGTCTGCTCTTCTCT 161102
111 CGGAGGGGGCGGTGAGTCTCCAGTATGTTGGCCCGAGGCTCTTGACCTTCTGGA 170
161103 CGGAGGGGGCGGTGAGTCTCCAGTATGTTGGCCCGAGGCTCTTGACCTTCTGGA 161162
171 CCNAGCGCCAGGACTGACGACGAGAGAGGGGGCTCACTCTTATCTCTGGGCAACC 230
161163 CCNAGCGCCAGGACTGACGACGAGAGAGGGGGCTCACTCTTATCTCTGGGCAACC 161222
231 ACTGACACAGAGGCGGCTCTCCCACTTAAATGATATACACATCACTGAGAGGGG 290
161223 ACTGACACAGAGGCGGCTCTCCCACTTAAATGATATACACATCACTGAGAGGGG 161282
291 ACCCAATCTGACTCTTCCCGGCTTGGAGCATTCGAGGCGCGGAGGAGAGTGCCTCCGCA 350
161283 ACCCAATCTGACTCTTCCCGGCTTGGAGCATTCGAGGCGCGGAGGAGAGTGCCTCCGCA 161342
351 GGCTTGGGCGGAGAGGCTCCAGAGAGGCACTGAGGCTGTGGCGGAGGCTCTGGAC 410
161343 GGCTTGGGCGGAGAGGCTCCAGAGAGGCACTGAGGCTGTGGCGGAGGCTCTGGAC 161402

```

Cy      411 ATCCGAGGACGACGAGGAAAGTCTCTCTGCGGCGCATCTGTAATAAACCCTTTTCTTT 470
Db      161403 ATCCGAGGACGACGAGGAAAGTCTCTCTGCGGCGCATCTGTAATAAACCCTTTTCTTT 161462
Cy      471 GTTTTAAAAA 482
Db      161463 GTTTTAAAAA 161474

RESULT 8
AC087790 171444 bp DNA linear HTG 09-MAY-2001
LOCUS Homo sapiens chromosome 15 clone RP11-276K9 map 15, WORKING DRAFT
DEFINITION
SEQUENCE 13 unordered pieces.
AC087790
AC087790.2 GI:13357344
HTG: HTG6_PHASE1; HTG6_DRAFT; HTG6_FULLTOP.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 171444)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 15, clone RP11-276K9
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 171444)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Bouckigalter,B., Brown,A.,
Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodgson,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,D.,
Gardner,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heathford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., Laroque,K., Lamazares,R., Landers,T.,
Lancosky,J., Levine,R., Liu,G., MacLean,C., MacDonald,P.,
Marquis,N., Mathews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheters,R., Meldrum,J., Meneus,L., Mihova,T., Mienga,Y.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Punkhang,P., Piere,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schupbach,R., Seaman,S., Severy,P.,
Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strausen,N., Subramanian,A., Talamas,J., Testa,S., Theodore,J.,
Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viet,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zaitoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-JAN-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 16, 2001 this sequence version replaced gi:12408501.
All repeats were identified using RepeatMasker:
Smit, A.P.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Project name: 112451
Center project name: 276_K_9

----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 165906 bases at least Q40
Consensus quality: 168587 bases at least Q30
Consensus quality: 169357 bases at least Q20
Insert size: 180000; agarose-IP
Insert size: 170244; sum-of-ctrls
Quality coverage: 5.7 in Q20 bases; agarose-IP
Quality coverage: 6.0 in Q20 bases; sum-of-ctrls

```

```

FEATURES
source
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
6092 6091: contig of 6091 bp in length
6192 6191: gap of 100 bp
6192 7986: contig of 1795 bp in length
7987 8086: gap of 100 bp
8086 8087: contig of 35853 bp in length
8087 43940: gap of 100 bp
43940 44039: gap of 100 bp
44039 45298: contig of 1259 bp in length
45298 45399: gap of 100 bp
45399 47595: contig of 2197 bp in length
47595 47695: gap of 100 bp
47695 52385: contig of 4690 bp in length
52385 52485: gap of 100 bp
52485 56310: contig of 3625 bp in length
56310 56410: gap of 100 bp
56410 71313: contig of 14903 bp in length
71313 71413: gap of 100 bp
71413 90993: contig of 13580 bp in length
90993 91093: gap of 100 bp
91093 111459: contig of 20366 bp in length
111459 111559: gap of 100 bp
111559 137726: contig of 26167 bp in length
137727 137826: gap of 100 bp
137827 170053: contig of 32227 bp in length
170054 170153: gap of 100 bp
170154 171444: contig of 1291 bp in length.
Location/Qualifiers
1. 171444
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-276K9"
/clone_11b="RP11-11 Human Male BAC"
1. 6091
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
6192..7986
/note="assembly_fragment"
8087..43939
/note="assembly_fragment"
44040..45298
/note="assembly_fragment"
45399..47595
/note="assembly_fragment"
47696..52385
/note="assembly_fragment"
52486..56310
/note="assembly_fragment"
56411..71313
/note="assembly_fragment"
71414..90993
/note="assembly_fragment"
91094..111459
/note="assembly_fragment"
111560..137726
/note="assembly_fragment"
137827..170053
/note="assembly_fragment"
170154..171444
/note="assembly_fragment"
clone_end:17
vector_side:right"

```


in the tiling path"
misc_feature 179373..181312

/note="overlap with CTD-2651B20, AC051619. This data is added to give overlap for long range contiguity."

ORIGIN

Query Match 80.0%; Score 398.4; DB 9; Length 181312;
Best Local Similarity 95.1%; Pred. No. 5.5e-88;
Matches 411; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 51 ACTGATACAGAGCCGCTTCTGAGTACAGCTGGCAACCGGCTCTGCTCTTCT 110
Db 168443 ACTGACATGCTCTCTCTTCTTCTGATCCACCGCAAGAGGCTCTGCTCTTCT 168402
QY 111 CGAGAGGCGCTGAGTGTCTCCAGTATGTTCCGACAGGCTCTTCCGACCTTCTGGA 170
Db 168403 CGAGAGGCGCTGAGTGTCTCCAGTATGTTCCGACAGGCTCTTCCGACCTTCTGGA 168462
QY 171 CCAAGCGCCAGAGAGTGCAGCCAGAGAGAGAGGAGGCTCACTTATCTCTGCGACCC 230
Db 168463 CCAAGCGCCAGAGAGTGCAGCCAGAGAGAGAGGAGGCTCACTTATCTCTGCGACCC 168522
QY 231 ACTGACAGAGAGGCGCTCTCCAGACTTAAATGATCAACACTTACCTGAGAGGG 290
Db 168523 ACTGACAGAGAGGCGCTCTCCAGACTTAAATGATCAACACTTACCTGAGAGGG 168582
QY 291 ACCCAATCTGAGACTCTTCCCGCTTGGGACATCGAGAGCGGGAAGCAGTGCAGCA 350
Db 168583 ACCCAATCTGAGACTCTTCCCGCTTGGGACATCGAGAGCGGGAAGCAGTGCAGCA 168642
QY 351 GGCTGGGCGAGAGAGCTCTCAGAGAGGAGGAGTCACTGAGGCTGCTGGGCGAGGCTCGAGC 410
Db 168643 GGCTGGGCGAGAGAGCTCTCAGAGAGGAGGAGTCACTGAGGCTGCTGGGCGAGGCTCGAGC 168702
QY 411 ATCCGACAGGACAGGAGAAATCTCTGAGGCGATCTGTAATAAATCTTTTCTTTT 470
Db 168703 ATCCGACAGGACAGGAGAAATCTCTGAGGCGATCTGTAATAAATCTTTTCTTTT 168762
QY 471 GTTTTTTAAAA 482
Db 168763 GTTTTTTAAAA 168774

RESULT 10

HS183A12F/C

LOCUS HS183A12F 217 bp DNA linear PRI 18-OCT-1995
DEFINITION H.sapiens CPG island DNA genomic MseI fragment, clone 183A12,
forward read cpg183A12.f1a.

ACCESSION

257552

VERSION

257552.1

KEYWORDS

CPG island; genomic MseI fragment.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

1

AUTHORS

Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.

TITLE

Purification of Cpg islands using a methylated DNA binding column

JOURNAL

Nat. Genet. 6 (3), 236-244 (1994)

MEDLINE

94282070

PUBMED

8012384

COMMENT

2 (bases 1 to 217)

REFERENCE

1

AUTHORS

Dodsworth,S.J., Huckle,E., Wilkinson,P. and Micklem,G.

TITLE

Direct Submission

Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, England. E-mail contact: humquerry@sanger.ac.uk

Vector: pGEM-5Zf(-)

Clones are available from the UK MRC Human Genome Mapping Project
Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:
http://www.hgmp.mrc.ac.uk/ for details
or contact: biohelp@hgmp.mrc.ac.uk.

Location/Qualifiers

1..217

source

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="183A12"
/sex="male"
/tissue_type="blood"
/clone_lib="CGI-1"
/dev_stage="adult"

ORIGIN

Query Match 43.6%; Score 217; DB 9; Length 217;
Best Local Similarity 100.0%; Pred. No. 4.4e-43;
Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 262 AATGATATACCACTAAGCTGTGAGAGGAGACCAATCTGACTCTTCCCGCTTGGGA 321
Db 217 AATGATATACCACTAAGCTGTGAGAGGAGACCAATCTGACTCTTCCCGCTTGGGA 158
QY 322 CATGAGAGCGGGAAGCAATGCGCCGACAGGCTTGGCCAGAGAGCTTCCAGAAAGGCA 381
Db 157 CATGAGAGCGGGAAGCAATGCGCCGACAGGCTTGGCCAGAGAGCTTCCAGAAAGGCA 98
QY 382 CTGAGAGCTGTGAGAGGAGGCTGCGACATCCGAGCAACAGGAAAGTCTCTGAGG 441
Db 97 CTGAGAGCTGTGAGAGGAGGCTGCGACATCCGAGCAACAGGAAAGTCTCTGAGG 38
QY 442 CGATCTGTAATTAACCTTTTCTTTTGTGTTTGA 478
Db 37 CGATCTGTAATTAACCTTTTCTTTTGTGTTTGA 1

RESULT 11

HS183A12R

LOCUS HS183A12R 218 bp DNA linear PRI 19-OCT-1995
DEFINITION H.sapiens CPG island DNA genomic MseI fragment, clone 183A12,
reverse read cpg183A12.r1a.

ACCESSION

259954

VERSION

259954.1

KEYWORDS

CPG island; genomic MseI fragment.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

1

AUTHORS

Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.

TITLE

Purification of Cpg islands using a methylated DNA binding column

JOURNAL

Nat. Genet. 6 (3), 236-244 (1994)

MEDLINE

94282070

PUBMED

8012384

COMMENT

2 (bases 1 to 218)

REFERENCE

1

AUTHORS

Dodsworth,S.J., Huckle,E., Wilkinson,P. and Micklem,G.

TITLE

Direct Submission

JOURNAL

Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, England. E-mail contact: humquerry@sanger.ac.uk

MEDLINE

94282070

PUBMED

8012384

COMMENT

2 (bases 1 to 218)

REFERENCE

1

AUTHORS

Dodsworth,S.J., Huckle,E., Wilkinson,P. and Micklem,G.

Vector: pGEM-5Zf(-)

Clones are available from the UK MRC Human Genome Mapping Project
Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:
http://www.hgmp.mrc.ac.uk/ for details
or contact: biohelp@hgmp.mrc.ac.uk.

Location/Qualifiers

1..218

source

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="183A12"
/sex="male"
/tissue_type="blood"
/clone_lib="CGI-1"
/dev_stage="adult"

ORIGIN

Query Match 40.1%; Score 199.8; DB 9; Length 218;
Best Local Similarity 97.7%; Pred. No. 8.1e-39;

TITLE
JOURNAL
MEDLINE
PUBMED
22388257
12477932
2 (bases 1 to 1354)
AUTHORS
REFERENCE
JOURNAL
TITLE
DIRECT SUBMISSION
Submitted (03-JUN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK
COMMENT
Contact: MGC help desk
Email: gcapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mcdpaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

FEATURES
SOURCE
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAC Plate: 58 Row: 9 Column: 17
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 13385245.
Location/Qualifiers
1..1354
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:35715 IMAGE:4989815"
/tissue_type="Colon, normal, 5 month old male mouse."
/clone_lib="NCI CGAP_C024"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1..1354
/gene="9030623N16R1K"
/db_xref="LocusID:66811"
/db_xref="MGI:1914061"
129..1091
/codon_start=1
/product="Riken CDNA 9030623N16"
/protein_id="AAH3111.1"
/db_xref="GI:2141408"
/db_xref="LocusID:66811"
/translation="MTAMDVLPFYPOPRHASPVLIVLVLASFLILPG
IRGSRMFWLRLVLSLFIAGEIVAVHSGDMFVGRVMTSYKAPSPRVGVHGLH
VGLAGVNTLGTGRCQINERTIDNERTKRLNDYTKVHVALEKIDPDTYLAEK
FTSPGCLVHQTDLAGHRAATLWVAFCEWITANLSPALVGLGALLTGAFIL
FGVFAFASISSVPLCHFRIGSAVLTPTVYGAFFWLTATGLSLLLGAGVILLHYRPS
ALRSLDLVSDCSNOKSGNSPLTINPPOHOLKSPDLNITILL"

gene
CDS

ORIGIN

Query Match 32.2%; Score 160.6; DB 10; Length 1354;
Best Local Similarity 65.8%; Pred. No. 4.4e-29;
Matches 331; Conservative 0; Mismatches 154; Indels 18; Gaps 6;

2 TCTAGCGTCCGCGTCTGCCCCGCTCCCGCTAGGCTCTCCGGCTCACCAGCTAGTACGA 61
810 TCCAGGATGCGCGCTGTG-CCACTTCGCGCTGGGCTCCGCGCTCCACGCTTACTAGC- 867
62 GCCCGCCTTCTGGGATCAGCTGGCAACCGCGCTGTGCTCTTCTCTGGAGGGCGG 121
868 GCGCCTCTTTTGGCTACAGCTGGCCACCGGATCTAGCTCTCTCTCGAGGGGGG 927
122 TGGTAGTCTCAGATATGTTCCGCCAGCGCTCTTTCGACCTTTTGAACCAAGCGCA 181
928 TGGTAGTCTTCCACTACTACTCGCCAGCGCCCTTCGCTCTTTCGATCTAAGTGTCA 987
182 AGGACT---GCAGCAGAGAGAGAGGGGGCTCACTCTTATCTCTGGCGACCACTGCACA 238
988 AAGACTGTAGCAACACAGGCTAAGAAATCACTCTCACTTCAACAACCGCAACAG 1047
239 AGCAGCGCGCTCTCCCAAGCTTAAATGATACCACTAAGCTGTAGGGGAGCCAAATC 298
1048 AACAGTGAAGAGTCCAGACTTAA---TATTAACAATCTCTGTGAAATAATGACTGC 1104
299 TGGACTCTTCCCGCGCTTGGGACATCGAGGCGGGAGAGAGTGCAGCCGCAAGCGCTG- 357
1105 CGGATTTCTACCCCTCTTTGGAGCCCATGACCTGAACTGTGTAAAGCGCTGCC 1164
358 GCCAGAGAGCTCCAGAAAGGCACTGAGCGCTGTGGCGAGGCGCTTCGACATCCGA 417
1165 AGAGGACATAGCTAGCTGTGAGAACTGTGTCCTGGGAGCAGCAAGAGGGGACCCCA 1224
418 GGCAC-----AGGAAAGTCTCTGGGGCCATCTGTAAATAAATCTTTTCTT 468
1225 TGGCCTTTCCTAGAGATGATTTCCAGAGAACTGTAAATAAATCTTTTGTTC 1284
469 TTGTTTAAATAAATAAATAA 491
1285 TTTTCTTCTTCAAAAAAAAAA 1307

RESULT 14
AL844566/C
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Mus musculus (house mouse)
HTG.
AL844566.8 GI:29823197
Mouse DNA sequence from clone RP23-173H17 on chromosome 2, complete
sequence.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 229583)
Beljan, S.
Direct Submission
Submitted (11-APR-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Apr 11, 2003 this sequence version replaced gi:23476718.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquerry@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Mp: MORNREP; Information on the MORNREP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormrep. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M3 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

RP23-173H17 is from the RPECI-23 Mouse BAC Library constructed by the group of Pieter de Jong.

For further details see <http://www.choxi.org/bacpac/home.htm>

VECTOR: pBACe3.6

Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to confirm this sequence. Sequence data from the whole genome shotgun alone has only been used where it has a phred quality of at least 30.

FEATURES

source

1..229583
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="2"
/clone="RP23-173H17"
/clone_1ib="RPECI-23"

ORIGIN

Query Match

Best Local Similarity 23.7%; Score 117.8; DB 10; Length 229583;

Matches 268; Conservative 0; Mismatches 157; Indels 16; Gaps 4;

60 GAGCGCGCGCTTCTGCGTACGCGTCAACCGCGCGCTTCTGCTCTTCTCGAGCGGCG 119
181832 GAGTCCCTCTGCTTCTTCCCTCCGCGTCAACCGCGCTTCTGCTCTTCTCGAGCGGCG 181773
120 CGTGTGAGTCTTCCAGTATGTTCCGCGCGCTTCTGCGCGCTTCTGCGCGCGCGCG 179
181772 GGTGTGATTCCTCACTATACCGCGCGCGCTTCTGCTCTTCTGAGTCTTAACTGT 181713
180 CAAGAGCT---GCAAGCAG 236
181712 CAAAGACTAGACACACAG 181653
237 CAAGCAGCGCGCTTCTCCAGACTTAAATGTAATCACTCACTAAGCTTGAAGGAGAGAG 236
181652 CGAACAGTTGAAGAGTCCAGACTTAA--TATTACCACTCTCTGTAATAATACTGAC 181596
297 TCTGAGACTCTTCCCGCGCTTGGGACATGCGAGCGCGGAGAGAGAGAGAGAGAGAGAGAG 356
181595 TCCGATTTCTACCTCTTCTTGGAGCCCATGAGACTTGAAGACTGTGTAAGAGCGGCTG 181536
357 GGCCAGAGAGAGTCCAG 415
181535 CCAGAGAGAGAGTCTGAGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 181476
416 CAGGAGAGCT---AGGAGAAAGTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 466
181475 CATGTGCTTGTCTGCTAG 181416
467 TTTTGTGTTTTTAAAAAAAT 487
181415 TTTTGTGTTTTTCTTCAAAAT 181395

RESULT 15

AF547266 5851 bp mRNA linear MAM 03-APR-2003
LOCUS AF547266
DEFINITION Sus scrofa dual oxidase 1 (Duox1) mRNA, complete cds.
ACCESSION AF547266
VERSION AF547266.1 GI:23664372

KEYWORDS

Sus scrofa (pig)
SOURCE Sus scrofa

ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.

REFERENCE

AUTHORS

1 (bases 1 to 5851)
Morand,S., Chaaroul,M., Kanlewski,J., Deme,D., Ohayon,R.,
Noel-Hudson,M.S., Virion,A. and Dupuy,C.

TITLE

Effect of iodide on nicotinamide adenine dinucleotide phosphate
oxidase activity and Duox2 protein expression in isolated porcine
thyroid follicles

Endocrinology 144 (4), 1241-1248 (2003)

JOURNAL

Endocrinology 144 (4), 1241-1248 (2003)

MEDLINE

22526370
12639906

PUBMED

2 (bases 1 to 5851)
Kantewski,J., Morand,S., Noel-Hudson,M.-S., Ohayon,R., Virion,A.
and Dupuy,C.

AUTHORS

Submitted (13-SEP-2002) Unite 486, INSERM University of Paris-Sud,
5 rue JB Clement, Chateauf-Malabry 92296, France

FEATURES

source

1..5851
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
1..5851
/gene="Duox1"
948..5609
/gene="Duox1"

gene

1..5851
/gene="Duox1"

CDS

1..5851
/note="NADPH:O2 oxidoreductase"

ORIGIN

1..5851
/note="NADPH:O2 oxidoreductase"

ORIGIN

1..5851
/note="NADPH:O2 oxidoreductase"

ORIGIN

1..5851
/note="NADPH:O2 oxidoreductase"

ORIGIN

1..5851
/note="NADPH:O2 oxidoreductase"

ORIGIN

1..5851
/note="NADPH:O2 oxidoreductase"

ORIGIN

1..5851
/note="NADPH:O2 oxidoreductase"

ORIGIN

1..5851
/note="NADPH:O2 oxidoreductase"

ORIGIN

1..5851
/note="NADPH:O2 oxidoreductase"

ORIGIN

1..5851
/note="NADPH:O2 oxidoreductase"

ORIGIN

1..5851
/note="NADPH:O2 oxidoreductase"

ORIGIN

1..5851
/note="NADPH:O2 oxidoreductase"

ORIGIN

1..5851
/note="NADPH:O2 oxidoreductase"

ORIGIN

1..5851
/note="NADPH:O2 oxidoreductase"

ORIGIN

1..5851
/note="NADPH:O2 oxidoreductase"

ORIGIN

1..5851
/note="NADPH:O2 oxidoreductase"

ORIGIN

1..5851
/note="NADPH:O2 oxidoreductase"

ORIGIN

1..5851
/note="NADPH:O2 oxidoreductase"

ORIGIN

1..5851
/note="NADPH:O2 oxidoreductase"

ORIGIN

1..5851
/note="NADPH:O2 oxidoreductase"

ORIGIN

1..5851
/note="NADPH:O2 oxidoreductase"

ORIGIN

1..5851
/note="NADPH:O2 oxidoreductase"

ORIGIN

1..5851
/note="NADPH:O2 oxidoreductase"

ORIGIN

1..5851
/note="NADPH:O2 oxidoreductase"

```
QY      1 CTCTAGGCTGCGCGCTCTGCGCGCTCTCCCGCTAGAGCTCTCCGCGCTCAACACTCACTAG 60
Db      694 CTCGAGGCTGCTCTCTTTGCGAGCT-CCGCTCGGCTCTCCGAGCTCAACACTCACTAG 752
QY      61 AGCGCGGCTTTCTGGGTCAGCTGAGCAACCGAGTCTGCTCTCTCTCTCTGAGAGGAGCC 120
Db      753 -GTGCTGCTTTTGATCAGCTGAGCAACCGAGTCTGCTCTCTCTCTCTGAGAGAGCG 811
QY      121 GTGCTGAGTCTCCAGTATGTTCCGCGCGCTCTCTGCGAGCCCTTCTGAGCCAAAGCGCC 180
Db      812 GTGCTGAGTCTCCAGTATGTTCCGCGCGCTCTCTGCGAGCCCTTCTGAGCCAAAGCGCC 871
QY      181 AAGGACTGCAAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 216
Db      872 AAGGACTGCAAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 907
```

Search completed: February 23, 2004, 21:57:54
Job time : 1969.03 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 19:33:52 ; Search time 201.357 Seconds

(without alignments)
10506.704 Million cell updates/sec

Title: US-09-936-456-3

Perfect score: 498

Sequence: 1 ctctacgcgtccgcctctgc.....aaaaaaaaataaagtcgacc 498

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 337363 segs, 212409941 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N Geneseg 29Jan04:*

1: Genesegm1980s:*

2: Genesegm1980s:*

3: Genesegm2000s:*

4: Genesegm2001as:*

5: Genesegm2001bs:*

6: Genesegm2002s:*

7: Genesegm2003as:*

8: Genesegm2003bs:*

9: Genesegm2003cs:*

10: Genesegm2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	498	100.0	498	AAA94624	AAA94624 Human CAS
2	465.4	93.5	1441	AAA94623	AAA94623 Human CAS
3	465.4	93.5	1474	ABL90535	ABL90535 Human pol
4	465.4	93.5	1491	ADD19228	ADD19228 Human CDN
5	453	91.0	1421	AAA96505	AAA96505 CDNA enco
6	421.8	84.7	1460	AAAC95564	AAAC95564 Human sec
7	421.8	84.7	1460	ABZ67235	ABZ67235 Human sec
8	421.8	84.7	1460	ABZ73640	ABZ73640 Secretd
9	421.8	84.7	1460	ADG20289	ADG20289 Human sec
10	398.4	80.0	2684	ADAC5627	ADAC5627 Human cod
11	398.4	80.0	5033	ABZ68115	ABZ68115 Human sec
12	398.4	80.0	5033	ABZ74587	ABZ74587 Secretd
13	398.4	80.0	5033	ADCC21005	ADCC21005 Human sec
14	46	9.2	3180	AAH17543	AAH17543 Human CDN
15	46	9.2	5049	AAI03901	AAI03901 Human rep
16	46	9.2	5049	AAI03900	AAI03900 Human rep
17	46	9.2	5049	AAI03901	AAI03901 Human rep
18	46	9.2	5049	AAI03900	AAI03900 Human rep
19	42	8.4	536	ADB88442	ADB88442 Xlnority
20	39	7.8	536	ADB88442	ADB88442 Xlnority
21	38.8	7.8	3786	AAI72683	AAI72683 DNA enco
22	38.8	7.8	8503	AAK81406	AAK81406 Human imm
23	38.6	7.8	2000	ADA71938	ADA71938 Rice gene

24	36.4	7.3	1751	6	ABL90392	ABL90392 Human pol
25	36.4	7.3	2074	4	AAI12572	AAI12572 Human pro
26	36.4	7.3	214	4	AAA93126	AAA93126 Human sec
27	36.4	7.3	2156	6	ABN59662	ABN59662 Novel hum
28	36.2	7.3	2776	4	AAK68475	AAK68475 Human imm
29	36.2	7.3	2776	4	AAK68476	AAK68476 Human imm
30	36.2	7.3	2776	4	AAI04043	AAI04043 Human rep
31	36.2	7.3	2776	4	AAI04045	AAI04045 Human rep
32	35.6	7.1	1810	3	AAZ56767	AAZ56767 Human tra
33	35.6	7.1	6232	6	AAI42135	AAI42135 Human kru
34	35.6	7.1	6232	6	AAI42134	AAI42134 Human kru
35	35.6	7.1	4798	6	ADB86070	ADB86070 Human cod
36	35.4	7.1	1464	6	ABQ99281	ABQ99281 Human cod
37	35.4	7.1	2080	2	AAZ10646	AAZ10646 Gene enco
38	35.4	7.1	2080	2	ADA55941	ADA55941 Gene enco
39	35.4	7.1	4200	3	AAA63349	AAA63349 Streptomy
40	35.4	7.1	6316	3	AAA63348	AAA63348 Streptomy
41	35.2	7.1	949	6	ABN74081	ABN74081 Bovine em
42	35.2	7.1	1395	7	AAI55820	AAI55820 Micromono
43	35.2	7.1	60136	7	AAI55810	AAI55810 Micromono
44	35	7.0	273	6	ABN18150	ABN18150 Human ORF
45	35	7.0	1395	4	AAI54132	AAI54132 Pseudomon

ALIGNMENTS

RESULT 1
ID AAA94624 standard; DNA; 498 BP.
XX AAA94624;
AC
XX
DT 11-JAN-2001 (first entry)
XX
DE Human CASB618 EST.
XX
KW Epitope; human; CASB618; chromosome 15; tumour; vaccine; cancer; ovarian; colon; autoimmune disease; HLA_A0201; expressed sequence tag; EST; ss.
XX
OS Homo sapiens.
XX
FN WO200053748-A2.
XX
PD 14-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-EP002048.
XX
PR 11-MAR-1999; 99GB-00005607.
XX
PR 01-SEP-1999; 99GB-00020590.
XX
PA (SMIX) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Bruck CEM, Cascart J, Coche T, Vinals Y De Bassolac;
XX
XX WPI; 2000-572268/53.
XX
DR New human CASB618 polypeptide, useful as a vaccine for prophylactic and therapeutic treatment of cancer, particularly ovarian or colon cancer, autoimmune diseases and related conditions.
XX
XX Claim 32; Page 62; 76pp; English.
XX
XX The present sequence is an expressed sequence tag (EST) for human CASB618 protein. The gene for human CASB618 is thought to be located on chromosome 15. CASB618 protein and epitopes of CASB618 protein (see AAH26327 to AAB26399) are useful in diagnosing the occurrence of tumour cells and in vaccines for prophylactic and therapeutic treatment of cancers, particularly ovarian or colon cancer, autoimmune diseases and related conditions.
XX
XX Sequence 498 BP; 97 A; 166 C; 137 G; 98 T; 0 U; 0 Other;

KM	vulnerans; anticonvulsant; antibacterial; antifungal; antiparasitic;
XN	cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW	neurological disease; infection; human; secreted protein; gene; ss.
OS	Homo sapiens.
XX	
PN	WC000190304-A2.
XX	
PD	29-NOV-2001.
XX	
PF	18-MAY-2001; 2001MO-US016450.
XX	
PR	19-MAY-2000; 2000US-0205515F.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
PI	Birse CE, Rosen CA;
WI	WPI; 2002-122018/16.
DR	P-PDB; ABB30126.
XX	
PT	Novel 1405 isolated polypeptides, useful for diagnosis, treatment and prevention of neural, immune system, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal and proliferative disorders.
PS	Claim 4; SEQ ID NO 1097; 2081pp + Sequence Listing; English.
CC	The invention relates to novel genes (ABL8944g-ABL9085) and proteins (ABB8904d-ABB9044d) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. CC The nucleic acids, proteins, antibodies and (anti)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, Rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d) wound healing CC infections. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPD at ftp.wipo.int/pub/published_pct_sequences
SQ	Sequence 1474 BP; 292 A; 480 C; 404 G; 296 T; 0 U; 2 Other:
	Query Match 93.5%; Score 465.4; DB 6; Length 1474; Best Local Similarity 99.4%; Pred. No. 9e-120; Matches 488; Conservative 0; Mismatches 1; Indels 2; Gaps 2
QY	1 CTCTAGCGTGCAGCTGTCGCCGCTCCCGCCTAGGCTCCTCGCGCTCACCACTCACTAGG 60
Db	965 CTCCTACGTCGCGCGCTGTGCCGCT-CCGCTTAGGCTCCTCGCGCTCACCACTCACTAG 1022
QY	61 AGGGCGGCTTTTGSGTCACTGTGCACAACGGCGTCTGTGTCTCTTCTCGAAGGGGCT 120
Db	1024 -GGCGCGCTTTTGSGTCACTGTGCACAACGGCGTCTGTGTCTCTTCTCGAAGGGGCT 1082
QY	121 GTGTAGTCTCCAGATATTGTGGCGCCAGCGCTCTTGCAACCCTTGCAACCAAAGCGCC 180
Db	1083 GTGTAGTCTCCAGATATTGTGGCGCCAGCGCTCTTGCAACCCTTGCAACCAAAGCGCC 1144
QY	181 AAGAAGTGAAGCCAGAGAGAGAGGGGCTCACTCTTAATCTCTGGCAACCACTGCAAG 240
Db	1143 AAGAAGTGAAGCCAGAGAGAGAGGGGCTCACTCTTAATCTCTGGCAACCACTGCAAG 1202
QY	241 CAGGCGGCTCTCCCAACTTAATAATGTATCAACTAACCTGTGAGGGGAGCCAAATCG 300
Db	1203 CAGGCGGCTCTCCCAACTTAATAATGTATCAACTAACCTGTGAGGGGAGCCAAATCG 1262
XY	301 GACTCTTCCCGCTTGGGACATCGAGGCCGGGAAACATGACCCGACGCTTGAGCC 360

Dd		1263 GACTCTTCCTCCGAGTTCGGACATCGAGGCCCGGAAGCATGTCCCGCAGGCCTGGGCC	1322
Qy		361 AGAGAAGCTCCAGAGAGGCGACTGAGCGCTGTCGGCGAGGCTTCGGACATCGCAAGC	420
Dd		1323 ACGAGAGCTCCAGAGAGGCGACTAGAGCGCTGTCGGCGAGGCTTCGGACATCCGCAAGC	1382
Qy		421 ACCAGGAGAAAGCTTCGGGGCGCATCTGTAAATAAACCTTTTTCCTTTGTTTTTAA	480
Dd		1383 ACCAGGAGAAAGCTTCGGGGCGCATCTGTAAATAAACCTTTTTCCTTTGTTTTTAA	1442
Qy		481 AAAAAATATAA 491	
Dd		1443 AAAAAAAAAA 1453	
RESULT 4			
ADDI9228	ID	ADDI9228 standard; cDNA, 1491 BP.	
AC	XX	ADDI9228;	
AD	XX	ADDI9228;	
DT	XX	15-JAN-2004 (first entry)	
Dd	XX	Human cDNA from secreted protein gene 45.	
Dd	XX	human secreted protein; cytosolic; antibacterial; virucide;	
Kw	XX	neuroprotective; gynaecological; gastrointestinal-gen; cardiant;	
Kw	XX	cardiovascular-gen; nephrotropic; antiinflammatory; muscular-gen;	
Kw	XX	respiratory-gen; immunosuppressive; cerebroprotective; viral infection;	
Kw	XX	nootropic; antiallergic; cancer; bacterial infection; blood disorder;	
Kw	XX	muscular disorder; immune system disorder; neurological disorder;	
Kw	XX	pulmonary disorder; cardiovascular disorder; renal disorder;	
Kw	XX	inflammatory disorder; proliferative disorder; human; ss; gene.	
Os	XX	Homo sapiens.	
Pn	XX	WO2003052377-A2.	
PD	XX	26-JUN-2003.	
Pf	XX	06-NOV-2002; 2002MO-USO35606.	
Pf	XX	07-NOV-2001; 2001US-0331046P.	
Pf	XX	(HUMA-) HUMAN GENOME SCI INC.	
Pt	XX	Rosen CA, Ruben SM;	
Pt	XX	WPI, 2003-533050/50.	
Pt	XX	P-PADB; ADDI9303.	
Pt	XX	New isolated nucleic acids encoding signal transduction pathway component	
Pt	XX	polypeptide, useful for diagnosing, treating, and/or preventing	
Pt	XX	disorders, such as cancer, infections, cardiovascular and inflammatory	
Pt	XX	diseases.	
Pt	XX	Claim 1; SEQ ID NO 55; 554bp; English.	
Pt	XX	The invention relates to an isolated nucleic acid molecule (cDNA)	
Pt	XX	encoding a human secreted protein, representing one of 85 novel genes.	
Pt	XX	Also included are recombinant vectors, host cells (expressing the	
Pt	XX	protein), the secreted proteins (including their fragments, epitopes and	
Pt	XX	homologues), an isolated antibody that binds specifically to the protein,	
Pt	XX	diagnosing a pathological condition or susceptibility to a pathological	
Pt	XX	condition (comprising determining the presence or absence of a mutation	
Pt	XX	in the nucleic acid and diagnosing a condition based on the presence or	
Pt	XX	absence of the mutation), diagnosing a pathological condition or	
Pt	XX	susceptibility to a pathological condition (comprising determining the	
Pt	XX	presence or amount of expression of the protein in a biological sample	
Pt	XX	and diagnosing a condition based on the presence or amount of expression	
Pt	XX	of the protein), preventing, treating or ameliorating a medical condition	

Qy	301	GACTCCTTCCCGCCCTTGGGACATCGAGAGCGCGGGAAGCAGTGCCTCCGACGCGCTGGGCGC	360
Db	1245	GACTCCTTCCCGCCCTTGGGAAATCGAGAGCGCGGGAAGCAGTGCCTCCGACGCGCTGGGCGC	1304
Qy	361	AGGAGAGCTCCAGAAAGGCACTGAGCGCTGCTGGCGGAGGCGCTCGGACATCCGAGGC	420
Db	1305	AGGAGAGCTCCAGAAAGGCACTGAGCGCGCTGCTGGCGGAGGCGCTCGGACATCCGAGGC	1364
Qy	421	ACCAGGGAAAGTCTCCTGGGGGAGTCTGTAAATAACCTTTTTCCTTTGTTTTT	477
Db	1365	ACCAGGGAAAGTCTCCTGGGGGAGTCTGTAAATAACCTTTTTCCTTTGTTTTT	1421
RESULT 6			
ID	AA095564	standard; cDNA; 1460 BP.	
XX	AA095564:		
XX	21-FEB-2001	(first entry)	
XX	Human secreted protein gene 44 SEQ ID NO:54.		
XX	Human; secreted protein; cytosolic; immunosuppressive; nocrotropic;		
XX	neuroprotective; antiviral; antiallergic; hepatotropic; antibacterial;		
XX	antiinflammatory; antitumor; vlnnary; anticonvulsant; antibacterial;		
XX	antifungal; antiparasitic; cardiac; cancer; immune disease; allergy;		
XX	cardiovascular disorder; wound healing; infection; neurological disease;		
XX	ss.		
XX	Homo sapiens.		
XX	MO200061596-A1.		
XX	19-OCT-2000.		
XX	06-APR-2000; 2000MO-US008993.		
XX	09-APR-1999; 99US-0128703P.		
XX	14-JAN-2000; 2000US-0176068B.		
XX	(HUMA-) HUMAN GENOME SCI INC.		
XX	(ROSE/) ROSEN C A.		
XX	Ruben SM, Komatsoulis G;		
XX	WPI: 2000-611865/58.		
XX	P-PSDB; AAB52055.		
XX	Fifty nucleic acid molecules encoding human secreted proteins, useful in		
XX	the prevention, treatment and diagnosis of cancer, immune disorders,		
XX	cardiovascular disorders and neurological diseases.		
XX	Claim 1; Page 443; 505pp; English.		
XX	Polynucleotide sequences AAC95521 - AAC95570 represent cDNA encoding 50		
XX	human secreted proteins AAB52012 - AAB52161. Sequences AAB52062 -		
XX	AAB52103 represent alternative polypeptides encoded by the genes, and		
XX	amino acid sequences with which they share homology. The genes and		
XX	proteins have activities dependent on the tissues and cells in which they		
XX	are expressed. Examples of their activities include cytostatic;		
XX	immunosuppressive; nocrotropic; neuroprotective; antiviral; anti-allergic;		
XX	hepatotropic; antibacterial; antiparasitic; antitumor; vlnnary;		
XX	anticonvulsant; antibacterial; antifungal; antiparasitic; and cardiac.		
XX	The secreted proteins, polypeptides, antagonists and agonists may be		
XX	useful in treating, preventing and/or diagnosing diseases and disorders		
XX	such as cancer, particularly breast and ovarian cancer, and other cancers		
XX	of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract,		
XX	liver, lung, or urogenital. Immune disorders such as Addison's disease,		
XX	allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,		
XX	diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid		
XX	arthritis and ulcerative colitis; cardiovascular disorders such as		
XX	myocardial ischemias; wound healing; neurological diseases such as		

Query Match	84.7%;	Score 421.8;	DB 3;	Length 1460;
Best Local Similarity	97.6%;	Pred. No. 1.4e-107;		
Matches 461;	Conservative 0;	Mismatches 7;	Indels 5;	Gaps 5;
QY	1	CTCTAGCGTGCAGCTCTGCGCTCCGCTCCGCTAGAGCTCTCGCGCTCACTCACTCACTAGAG	60	
Db	960	CTCTAGCGTGCAGCTCTGCGCTCCGCTCCGCTAGAGCTCTCGCGCTCACTCACTCACTAGAG	1018	
QY	61	AGCGCGCGCTTCTGGGTACAGCTTGAGACCGCGCGCTCTGTGCTCTTCTCGAGAGGGGC	120	
Db	1019	-GGCGCGCTTCTGGGTACAGCTTGAGACCGCGCGCTCTGTGCTCTTCTCGAGAGGGGC	1077	
QY	121	GTGATGAGCTCTCAATATGTTTGAGCCAGCGCTCTTCTGCAACCTTCTGAGCCAAAGCGCC	180	
Db	1078	GTGATGAGCTCTCAATATGTTTGAGCCAGCGCTCTTCTGCAACCTTCTGAGCCAAAGCGCC	1137	
QY	181	AAGGACTGC-AGCCAGAGAGAGGGGGCTCACTCTTATCTCTGCGGACCCACTGACAA	239	
Db	1138	AAGGACTGCAGAGCCAGAGAGAGGGGGCTCACTCTTATCTCTGAGGACCCACTGACAA	1196	
QY	240	GCAGGCGCGCTCTCCAGAGC-TTAAATAGTATCACTCACTAACTGTAGAGGGAGCCAAATC	298	
Db	1197	GCAGGCGCGCTCTCCAGAGCTTTAAATAGTATCACTCACTAACTGTAGAGGGAGCCAAATC	1256	
QY	299	TGAGACTCTCTCCCGCTCTGGGACATCGCAGGCCCGGAGACAGTGCCTCGCAGGCTTGCG	358	
Db	1257	TGAGACTCTCTCCCGCTCTGGGACATCGCAGGCCCGGAGACAGTGCCTCGCAGGCTTGCG	1314	
QY	359	CCAGGAGACTCCAGAGAGGGCATTGAGCGCTGTGCGGAGAGCGCTCGGACATCGGAG	418	
Db	1317	CCAGGAGACTCCAGAGAGGGCATTGAGCGCTGTGCGGAGAGCGCTCGGACATCGGAG	1378	
QY	419	GCACACGAGAAAGTCTCTGGGAGCATCTGTAATAAACCTTTTCTTTTGTGTTTAA	478	
Db	1377	GCACACGAGAAAGTCTCTGGGAGCATCTGTAATAAACCTTTTCTTTTGTGTTTAA	1438	
QY	479	AAAAAAAAATAAA 491		
Db	1437	AAAAAAAAAAAAA 1449		

RESULT 7
 AB267235
 ID AB267235 standard; cDNA; 1460 BP.
 XX
 XX AC
 XX AB267235;
 DT 26-MAR-2003 (first entry)
 XX
 XX Human secreted protein encoding cDNA SEQ ID NO 355.
 XX
 XX Human; secreted protein; neutrotropic; neuroprotective; cytostatic;
 KW virucide; dermatological; immunosuppressive; antiinflammatory; anti-HIV;
 KW vulnery; antibacterial; antiparkinsonian; antischling; antinaemic;
 KW antiasthmatic; cancer; antihypertensive; hepatotropic; cerebroprotective;
 KW antiinflammatory; antiallergic; antidiabetic; antidiuretic; anticonvulsant;
 KW antifungal; antiparasitic; cardiac; immune disorder; infection; vaccine;
 KW cardiovascular disorder; neurological disease; nephrotropic;
 KW gene therapy; gene; chromosome 9p21; ds.
 OS Homo sapiens.
 XX
 XX WO200277186-A2.
 XX
 XX 03-OCT-2002.
 PD

XX 26-MAR-2002; 2002WO-US009188.
 XX 27-MAR-2001; 2001US-0278650P.
 PR 12-SEP-2001; 2001US-00950082.
 PR 12-SEP-2001; 2001US-00950083.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Ruben SM;
 XX PI
 DR WPI: 2003-040583/03.
 DR P-PSDB; ABP99814.
 XX
 PT New human secreted proteins encoded by genes contained in cDNA clones
 PT (e.g. HOCAC19), useful for preventing, treating or diagnosing e.g. AIDS,
 PT multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or
 PT West Nile fever.
 XX
 PS Claim 7; Page 1353; 2423pp; English.
 XX
 CC The invention relates to novel human genes (AB266891-AB268209) and the
 CC encoded secreted proteins (ABP99470-ABP99872) useful for preventing,
 CC treating or ameliorating medical conditions e.g. by protein or gene
 CC therapy. The genes are isolated from a range of human tissues disclosed
 CC in the specification. The nucleic acids, proteins, antibodies and
 CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
 CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the
 CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
 CC lung or urogenital; (b) immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
 CC myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.
 CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
 CC bacterial, fungal and parasitic infections
 XX
 SQ Sequence 1460 BP; 292 A; 472 C; 398 G; 298 T; 0 U; 0 Other;
 Query Match 84.7%; Score 421.8; DB 7; Length 1460;
 Best Local Similarity 97.6%; Pred No. 14e-107;
 Matches 461; Conservative 0; Mismatches 7; Indels 5; Gaps 5;
 QY 1 CTCCTAGCGTGGCGCTCTGCGCGCTCCGCTAGGCTCTCCGCGCTACACACTAGTACG 60
 DB 960 CTCCTAGCGTGGCGCTCTGCGCGCTCCGCTAGGCTCTCCGCGCTACACACTAGTACG 1018
 QY 61 AAGCGCGCTTCTGGGTACAGCTGGCAACCGCGCTCTGCTCTCTCTCTGAGGGGCC 120
 DB 1019 -GGCGCGCTTCTGGGTACAGCTGGCAACCGCGCTCTGCTCTCTCTCTGAGGGGCC 1077
 QY 121 GTGGTAGTCTCCAGTATGTTGGCCCAAGCGCTCTCCAGCTCTTGGACCAAGAGGCC 180
 DB 1078 GTGGTAGTCTCCAGTATGTTGGCCCAAGCGCTCTCCAGCTCTTGGACCAAGAGGCC 1137
 QY 181 AAGGACTGC-AAGCAGAGAGAGAGGGGGCTCACCTTTATCTCTGGCGAACCACTGCAAA 239
 DB 1138 AAGGACTGCAGCAGAGAGAGAGGGGGGGCTCACCTTTATCT-GGGAGCCACAGCAAA 1196
 QY 240 GCAGGCGCTCTCCAGAC-TTAAATGATACACACTACACTGAGAGAGAGAGCCAAATC 298
 DB 1197 GCAGGCGCTCTCCAGAC-TTAAATGATACACACTACACTGAGAGAGAGAGCCAAATC 1256
 QY 299 TGAGTCTCTTCCCGCTTGGAGCATTCGAGGCGCGGAAAGAGTGCCTCCGCAAGGCTGGG 358
 DB 1257 TGAGTCTCTTCCCGCTTGGAGCATTCGAGGCGCGGAAAGAGTGCCTCCGCAAGGCTGGG 1316
 QY 359 CCAGAGAGCTCCAG 418
 DB 1317 CCAGAGAGAGCTCCAG 1376
 QY 419 GCACGAGGAAAGTCTCTGAGGCGAGTCTGTAATTAACCTTTTCTTTGTTTTTA 478

DB 1377 GCACGAGGAAAGTCTCTGAGGCGAGTCTGTAATTAACCTTTTCTTTGTTTTTA 1436
 QY 479 AAAAAAAAAA 491
 DB 1437 AAAAAAAAAA 1449
 RESULT 8
 ID AB273640
 AC AB273640;
 XX
 DT 12-MAY-2003 (first entry)
 DE
 XX Secreted protein-encoding gene 360 cDNA clone HUPCL21, SEQ ID NO:370.
 XX
 KW Human; secreted protein; cancer; tumour; hyperproliferative disorder;
 KW autoimmune disorder; inflammation; angiogenic diseases; AIDS;
 KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
 KW drug screening; chromosome identification; chromosome mapping;
 KW cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;
 KW antianemic; vulnery; chromosome 3p21; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200277013-A2.
 PD 03-OCT-2002.
 XX
 PF 26-MAR-2002; 2002WO-US009370.
 PR 27-MAR-2001; 2001US-0278650P.
 PR 12-SEP-2001; 2001US-00950082.
 PR 12-SEP-2001; 2001US-00950083.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI: 2003-040578/03.
 DR P-PSDB; ABP01306.
 XX
 PT New human secreted proteins and nucleic acids, useful for detecting or
 PT treating cancer or other hyperproliferative disorders, autoimmune
 PT disorders, inflammatory disorders, HIV disease, hepatitis or anaemia.
 XX
 PS Claim 21; Page 1345; 2474pp; English.
 XX
 CC AB273281-AB273697 represent cDNAs corresponding to 391 human secreted
 CC protein genes, and ABP00947-ABP01363 represent the proteins they encode.
 CC AB273698-AB274687 represent human secreted protein genomic fragments. The
 CC invention also encompasses antibodies specific for the secreted proteins,
 CC the use of the secreted proteins in drug screening and recombinant
 CC vectors and host cells comprising a nucleic acid of the invention. The
 CC secreted proteins are thought to be involved in biological activities
 CC associated with cellular signalling, cellular differentiation, cell
 CC migration, proinflammatory activation and neurotransmitter activity. The
 CC secreted proteins, nucleic acids encoding them, antibodies or antibody
 CC fragments specific for the secreted proteins, and modulators of protein
 CC activity are useful for diagnosing or treating cancers or other
 CC hyperproliferative disorders. Additionally, the secreted proteins and
 CC their nucleic acids may also be used in the treatment of autoimmune
 CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
 CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
 CC wound healing. Nucleic acids of the invention may be used for chromosome
 CC identification, chromosome mapping, in gene therapy, for identifying
 CC individuals from minute biological samples, as hybridisation probes, and
 CC as molecular weight markers. The present sequence represents a human
 CC secreted protein-encoding cDNA clone of the invention.
 XX
 SQ Sequence 1460 BP; 292 A; 472 C; 398 G; 298 T; 0 U; 0 Other;

Pt	Rosen CA, Ruben SM;
DR	WP1; 2003-129287/12.
Pt	New human secreted proteins and nucleic acid molecules, useful for preparing a diagnostic or pharmaceutical composition for diagnosing, preventing or treating hematopoietic or hematologic disorders, e.g. anemia or hemophilia.
Pt	Claim 1; SEQ ID NO 238; 1512bp; English.
Pt	The invention comprises the amino acid and coding sequences of human secreted proteins. The DNA and protein sequences of the invention are useful for detecting, preventing, diagnosing, prognosticating, treating or ameliorating: haematopoietic or hematological disorders (e.g. anaemia and haemophilia); inflammatory disorders (e.g. inflammatory bowel disease and Crohn's disease); neoplastic disease (e.g. cancer and Leukaemia); wound healing and disorders of epithelial cell proliferation; immune disorders (e.g. autoimmune disorders and asthmatic disorders); cardiovascular disorders (e.g. atherosclerosis and myocarditis); infectious disease (e.g. HIV/AIDS); endocrine disorders (e.g. diabetes); and gastrointestinal disorders (e.g. duodenal ulcers and gastroenteritis). The present DNA sequence encodes a human secreted protein of the invention.
SQ	Sequence 1460 BP; 292 A; 472 C; 398 G; 298 T; 0 U; 0 Other;
Query March	84.7%; Score 421.8; DB 9; Length 1460;
Best Local Similarity	97.6%; Pred. No. 1.4e-107;
Matches 481; Conservative	0; Mismatches 7; Indels 5; Gaps 5;
Dn	CTCTAAGCGTGCAGCTGTGCGCCGCTCCGGCTTCCTCCGCTACACTCAATGATG 60
Yy	1 CTCTAAGCGTGCAGCTGTGCGCCGCTCCGGCTTCCTCCGCTACACTCAATGATG 60
Dn	960 CTCCTAAGCGTGCAGCTGTGCGCCGCTCCGGCTTCCTCCGCTACACTCAATGATG 1018
Yy	61 AGGCGGCCCTTTTGSGGTACAGCTGGCAAACGGCGCTCTGTCCTTTCTGGAGGGGCG 120
Dn	1019 -GGCGCGCCTTTTGSGGTACAGCTGGCAAACGGCGCTCTGTCCTTTCTCCAGGGGCG 1077
Yy	121 GTGATGATCTCCAGTATGTTCCGGCCAGCGCTCTTGCAACCCTTTGGACCAAGCGCC 180
Dn	1078 GTGATGATCTCCAGTATGTTCCGGCCAGCGCTCTTGCAACCCTTTGGACCAAGCGCC 1137
Yy	181 AAGAGATCG-AAGCAGAGAGAAGGGGGCTCACCTCTTATCTTGAGGAGCCACTGCAACA 239
Dn	1138 AAGAGATCGAAGCAGAGAGAAGGGGGGCTCACCTCTTATCT-TGGAGAACCACTGCAACA 1156
Yy	240 GCAGGCGCGCTCTCCAGAC-TTAAATATGATACACTTAACCTGTAGAGGGGAGCCCAATC 298
Dn	1197 GCAGGCGCGCTCTCCAGACTTTAAATATGATACACTTAACCTGTAGAGGGGAGCCCAATC 1256
Yy	299 TGGAATCTTCCCCTTTGGGACATGCGAGCCGGGAGACAGTGCCTCCAGGCTGGG 358
Dn	1257 TGGAATCTTCCCCTTTGGGACATGCGAGCCGGGAGAACAGTGCCTCCAGGCTGGG 1316
Yy	359 CGAGGAGAGTCCAGGAGAGGGGACATGAGCGCTGTGGCGGAGGCTCGGACATCGCAG 418
Dn	1317 CGAGGAGAGTCCAGGAGAGGGGACATGAGCGCTGTGGCGGAGGCTCGGACATCGCAG 1378
Yy	419 GCACCAAGGAAAGTCTCTCGGGGCGATCTGTAAATAAACTTTTTTTCTTTGTTTTTA 478
Dn	1377 GCACCAAGGAAAGTCTCTCGGGGCGATCTGTAAATAAACTTTTTTTCTTTGTTTTTA 1436
Yy	479 AAAAAAAAAAAAAA 491
Dn	1437 AAAAAAAAAAAAAA 1449

RESULT 10
ID ADA53627/c
ADA53627 standard; cDNA; 2684 BP.
XX

AC	ADAS3627.
XX	
DT	20-NOV-2003 (first entry)
XX	
DE	Human coding sequence, SEQ ID 1195.
XX	
KW	Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
XX	Gene Therapy; human; secretory protein; membrane proteins; cancer;
KM	Inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX	
OS	Homo sapiens.
FN	EP1293569-A2.
PD	19-MAR-2003.
XX	
PP	21-MAR-2002; 2002EP-00006586.
XX	
PR	14-SEP-2001; 2001JP-00328381.
PR	24-JAN-2002; 2002US-0350435P.
XX	
PA	(HELI-) HELIX RES INST.
XX	(REAS-) RES ASSOC BIOTECHNOLOGY.
P1	Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,
P1	Yamamoto U, Isono Y, Hiro Y, Otsuka K, Nagai K, Irie R, Tanachika I,
P1	Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y,
DR	WP1; 2003-395539/38.
XX	P-PSDB; ADAS5266.
PT	New polynucleotides encoding full-length polypeptides, e.g. secretory
PT	and/or membrane proteins, useful for developing medicines for diseases in
PT	which the gene is involved, or as target molecules for gene therapy.
XX	
PS	Claim 1; SEQ ID NO 1195; 2055P; English.
CC	The present invention relates to novel human secretory or membrane
CC	proteins (ADAS4072-ADAS5710) and their coding sequences (ADAS2433-
CC	ADAS4071). The coding sequences are useful in the gene therapy of
CC	diseases caused by abnormalities of the proteins, e.g. cancer,
CC	inflammatory diseases, osteoporosis or neurological disease.
XX	
SS	Sequence 2684 BP; 599 A; 785 C; 684 G; 616 T; 0 U; 0 Other;
Query Match	80.0%; Score 398.4; DB 7; Length 2684;
Best Local Similarity	95.1%; Pred. No. 6,1e-101;
Matches 411; Conservative	0; Mismatches 21; Indels 0; Gaps 0;
QY	51 ACTCAGTACGAGCGCCGCTTCGTGGATCAAGCTGGCAACGGGCGTCTGTGCTTTCT
DB	2383 ACTCCATCATGCCTCTCTTTCTTGATCCCAACGCCAAGGCGTCTGTGCTTTCT
QY	111 CGAGGGGGCCGTGTAGTCTCCAGTATGTTGGCCAGGCGTCTTGACACTTTGTGA
DB	2323 CGAGGGGGCCGTGTAGTCTCCAGTATGTTGGCCAGGCGTCTTGACACTTTGTGA
QY	171 CCAAAGCGCAAGACTGACGACGACGAGAGAGGGGGCTCACTTTATTCTCGGACACC
DB	2263 CCAAAGCGCAAGACTGACGACGAGAGAGGGGGCTCACTTTATTCTCGGACACC
QY	231 ACTGCAACAACGAGCGCGTCTCCAGACTTAATAATGATACCACTTAAGTGGAGGG
DB	2203 ACTGCAACAACGAGCGCGTCTCCAGACTTAATAATGATACCACTTAAGTGGAGGG
QY	291 ACCCAATCTGACTCTTCCTCCGCGCTGGGACATGACAGGCGGGAAGCACTGCGCCA
DB	2143 ACCCAATCTGACTCTTCCTCCGCGCTGGGACATGACAGGCGGGAAGCACTGCGCCA
QY	351 GGCTTGAGCCAGAGAGCTTCAGAAAGGCACTGAGCGCTCTGGCGGAGGCTTGAGC
DB	2083 GGCTTGAGCCAGAGAGCTTCAGAAAGGCACTGAGCGCTCTGGCGGAGGCTTGAGC

Oy	411	ATCCGAGGACACCGAGGAAACTCTCCGCGGGCCATCTGTAATAAACCTTTTCTTT	470
Db	2023	ATCCGAGGACACCGAGGAAAGCTCTCTGGGGCGATCTGTAAATTAACCTTTTCTTTT	1964
Oy	471	GTTTTAAAAA 482	
Db	1963	GTTTTAAAAA 1952	
RESULT 11			
ID	ABZ68115		
XX	ABZ68115	standard; DNA; 5033 BP.	
AC	ABZ68115;		
XX			
DT	26-MAR-2003	(first entry)	
DE			
XX	Human secreted protein encoding genomic DNA SEQ ID NO 1638.		
KM	Human; secreted protein; nontropic; neuroprotective; cytostatic;		
KM	vincide; dermatological; immunosuppressive; antiinflammatory; anti-HIV;		
KM	vulnery; antibacterial; antiparkinsonian; antischlicking; antianemic;		
KM	antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective;		
KM	antiflammatory; antiallergic; antidiabetic; antitumor; anticonvulsant;		
KM	antifungal; antiparasitic; cardiac; immune disorder; infection; vaccine;		
KM	cardiovascular disorder; neurological disease; nephrotropic;		
KM	gene therapy; gene; ds.		
XX			
OS	Homo sapiens.		
PN	MO200277186-A2.		
XX			
PD	03-OCT-2002.		
PE	26-MAR-2002; 2002W0-US009188.		
XX			
PR	27-MAR-2001; 2001US-0278650P.		
PR	12-SEP-2001; 2001US-00950082.		
XX	12-SEP-2001; 2001US-00950083.		
PA	(HUMA-) HUMAN GENOME SCI INC.		
PI	Rosen CA, Ruben SM;		
XX	WPI; 2003-040583/03.		
PT	New human secreted proteins encoded by genes contained in cDNA clones		
PT	(e.g. HGCGC9), useful for preventing, treating or diagnosing e.g. AIDS,		
PT	multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or		
PT	West Nile fever.		
XX			
PS	Disclosure; Page 2263-2264; 2423pp; English.		
XX			
CC	The invention relates to novel human genes (ABZ68115-ABZ68209) and the		
CC	encoded secreted proteins (ABP99470-ABP99872) useful for preventing,		
CC	treating or ameliorating medical conditions e.g. by protein or gene		
CC	therapy. The genes are isolated from a range of human tissues disclosed		
CC	(ant)agonists are useful in the diagnosis, treatment and prevention of:		
CC	(a) cancer, e.g. breast and ovarian cancer and other cancers of the		
CC	adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,		
CC	lung or urogenital; (b) immune disorders e.g. Addison's disease,		
CC	allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,		
CC	diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid		
CC	arthritis and ulcerative colitis; (c) cardiovascular disorders such as		
CC	myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.		
CC	cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,		
CC	bacterial, fungal and parasitic infections		
XX			
SO	Sequence 5033 BP; 1061 A; 1409 C; 1465 G; 1098 T; 0 U; 0 Other;		
Query Match	80.0%; Score 398.4; DB 7; Length 5033;		
Best Local Similarity	95.1%; Pred.No. 7.9e-101;		

	Matches	411; Conservative	0; Mismatches	21; Indels	0; Gaps	0;
QY	51	ACTCAGTACGAGCGCCGCGCTTCTGGGATCAGCTGGCAACCGGGCTCTGTGCTCTTCT	110			
Db	3340	ACTCCACATGAGCCCTCTTCTTTCATCCCAACCGCCACAGGGGCTCTGTGCTCTTCT	3399			
QY	111	CGAGAGGGGCGGTGTGATGCTCCAGTATGTTGGCCAGCGCTCTTGGACCTTCTGGA	170			
Db	3400	CGAGAGGGGCGGTGTGATGCTCCAGTATGTTGGCCAGCGCTCTTGGACCTTCTGGA	3459			
QY	171	CCAAAGGCCCAAGAGACTGCAAGCAGAGAGAAGG999CTCACTTCTTATCTTGGCGACCC	230			
Db	3460	CCAAAGGCCCAAGAGACTGCAAGCAGAGAGAAGG999CTCACTTCTTATCTTGGCGACCC	3519			
QY	231	ACTGCAACAAGCAGGCGGCTCTCCCAAGTTAAATGTATACCACTAACCTGTGAAGGGG	290			
Db	3520	ACTGCAACAAGCAGGCGGCTCTCTCCAGATTAAATGTATACCACTAACCTGTGAAGGGG	3579			
QY	291	ACCCAACTGGAATCTCTTCCCGCCCTTGGGACATCGAGGCGG9GAAGCATGGCCGCA	350			
Db	3580	ACCCAACTGGAATCTCTTCCCGCCCTTGGGACATCGAGGCGG9GAAGCATGGCCGCA	3639			
QY	351	GGGCTGGGCCAAGAGAGCTCCAGGAAGGGCACTAAGCGCTGTGGGCCAGG9CTTCGAC	410			
Db	3640	GGGCTGGGCCAAGAGAGCTCCAGGAAGGGCACTAAGCGCTGTGGGCCAGG9CTTCGAC	3699			
QY	411	ATCCGAGAGGACCAAGGAAAGTCTCTGGGCGCATGTGAATAAACCTTTTCTTTT	470			
Db	3700	ATCCGAGAGGACCAAGGAAAGTCTCTCTGGGCGCATGTGAATAAACCTTTTCTTTT	3759			
QY	471	GTTTTTTAAAAA 482				
Db	3760	GTTTTTTAAAAA 3771				
RESULT 12						
ABZ74587	ABZ74587 standard; DNA; 5033 bp.					
XX	ABZ74587;					
XX	12-MAY-2003 (first entry)					
DB	Secreted protein gene 360 genomic fragment HUFCL31, SEQ ID NO:1734.					
XX	Human; secreted protein; cancer; tumour; hyperproliferative disorder;					
KW	autoimmune disorder; inflammation; angiogenic diseases; AIDS;					
KW	acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;					
KW	drug screening; chromosome identification; chromosome mapping;					
KW	cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;					
KW	antianaemic; vulnerary; chromosome 9p21; gene; ds.					
XX	Homo sapiens.					
OS	MO200277013-A2.					
PN	03-OCT-2002.					
PD	26-MAR-2002; 2002MO-US009370.					
PF	27-MAR-2001; 2001US-0278650P.					
XX	PR 12-SEP-2001; 2001US-00950082.					
PR	PR 12-SEP-2001; 2001US-00950083.					
XX	(HUMA-) HUMAN GENOME SCI INC.					
PA	Rosen CA, Ruben SM;					
XX	PI					
XX	DR					
XX	WPI, 2003-040578/03.					
PT	New human secreted proteins and nucleic acids, useful for detecting or					
PT	treating cancer or other hyperproliferative disorders, autoimmune					
PT	disorders, inflammatory disorders, HIV disease, hepatitis or anemia.					

PS Disclosure; Page 2315-2316; 2474bp; English.

CC AB273281-AB273697 represent cDNAs corresponding to 391 human secreted
XX protein genes, and ABP00947-ABP01361 represent the proteins they encode.
CC AB273698-AB274687 represent human secreted protein genomic fragments. The
CC invention also encompasses antibodies specific for the secreted proteins,
CC the use of the secreted proteins in drug screening and recombinant
CC vectors and host cells comprising a nucleic acid of the invention. The
CC secreted proteins are thought to be involved in biological activities
CC associated with cellular signalling, cellular differentiation, cell
CC migration, pro-hormone activation and neurotransmitter activity. The
CC secreted proteins, nucleic acids encoding them, antibodies or antibody
CC fragments specific for the secreted proteins, and modulators of protein
CC activity are useful for diagnosing or treating cancers or other
CC hyperproliferative disorders. Additionally, the secreted proteins and
CC their nucleic acids may also be used in the treatment of autoimmune
CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
CC wound healing. Nucleic acids of the invention may be used for chromosome
CC identification, chromosome mapping, in gene therapy, for identifying
CC individuals from minute biological samples, as hybridisation probes, and
CC as molecular weight markers. The present sequence represents a human
CC secreted protein genomic fragment referred to in the disclosure of the
XX invention

SQ Sequence 5033 BP; 1061 A; 1409 C; 1465 G; 1098 T; 0 U; 0 Other;

Query Match 80.0%; Score 398.4; DB 7; Length 5033;
Best Local Similarity 95.1%; Pred. No. 7.9e-101;
Matches 411; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

OY 51 ACTCGATGAGCAGCGCGCTTCTTGAGTCACGCCTGGCAACCGGCGCTGTGCTCTTCT 110
DB 3340 ACTCCAGCATGCCCTCTTTCTTGATGCCACCAGCGAAGGGGCTGTGCTCTTCT 3399

OY 111 CGAGAGGGCGCTGTGAATCTCCAGTAATTTCGGCCCAAGCGCTTTGGACCTTCAG 170
DB 3400 CGAGAGGGCGCTGTGAATCTCCAGTAATTTCGGCCCAAGGGGCTTTGGACCTTCAG 3459

OY 171 CCAGAACCGCAAGACTCACACAGAGAAGAGGGGGCTCACTTATCTTCGGGACCC 230
DB 3460 CCAGAACCGCAAGACTCACACAGAGAAGAGGGGGCTCACTTATCTTCGGGACCC 3519

OY 231 ACTGCACAAGCAGSCCGCTTCTCCAGACTTAATGTATACCACTTAACCTGTAGGGG 290
DB 3520 ACTGCACAAGCAGSCCGCTTCTCCAGACTTAATGTATACCACTTAACCTGTAGGGG 3579

OY 291 ACCCATCTCGAGCTCTTCCCAGCTTGGGACATCGAAGCGGGAAGAGAGCCGCCA 350
DB 3580 ACCCATCTCGAGCTCTTCCCAGCTTGGGACATCGAAGCGGGAAGAGAGGCCGCCA 3639

OY 351 GGCGTTGGGCAGAGAGCTCCAGAAAGGCACTGAAGCGTGGCGGCGAGGCTTCGAC 410
DB 3640 GGCGTTGGGCAGAGAGCTCCAGAAAGGCACTGAAGCGTGGCGGCGAGGCTTCGAC 3699

OY 411 ATCCGAGGCAACAAGAAAATCTCCCTGGGGCAGTCTGTAATAAACCCTTTTCTTT 470
DB 3700 ATCCGAGGCAACAAGAAAATCTCCCTGGGGCAGTCTGTAATAAACCCTTTTCTTT 3759

OY 471 GTTTTTTAAAAA 482
DB 3760 GTTTTTTAAAAA 3771

RESULT 13
ADC21005
ID ADC21005 standard; DNA; 5033 BP.
XX AC ADC21005;
XX DT 18-DEC-2003 (first entry)
XX XX

CC CDNA's easily without any specialised methods, AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AB03446 to AB05893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 3180 BP; 687 A; 881 C; 924 G; 688 T; 0 U; 0 Other;
Query Match 9.2%; Score 46; DB 4; Length 3180;
Best Local Similarity 57.7%; Pred. No. 0.022;
Matches 82; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 91 GGGGTCCTGATGCTTCTCTCTCGAGGAGGCGCGTGTGAGTCTCCAGTATGTCGAGCCAGC 150
Db 1564 GGGGTTGTATCTCGGAGTAGGGGAGACAGAGTTGGGCTCAGAGATGCTCGACAGC 1505
QY 151 GCTCTTGGACACCTTCTTGACCAAGCGCCAGAGATGTCAGCCAGAGAGAGGGGCTCA 210
Db 1504 AGGGTCTCAGCCTGTGGGCGCCATGTGATTAAGCACTGGAGGCCGAGATAGGGGCTTT 1445
QY 211 CTTCTTATCTCTGCGGCGACCCAC 232
Db 1444 AGAGTAGCCCGCTGGAGGCTC 1423
RESULT 15
AAL03901/c
ID AAL03901 standard; DNA; 5049 BP.
XX AAL03901;
AC
XX 21-NOV-2001 (first entry)
DE Human reproductive system related antigen DNA SEQ ID NO: 6589.
XX Human reproductive system related antigen; reproductive system disorder;
KW Human; gene therapy; ds.
XX Homo sapiens.
OS
XX
XX WO200155320-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001339.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198173P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217467P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225577P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226799P.
PR 22-AUG-2000; 2000US-0226811P.
PR 22-AUG-2000; 2000US-0226868P.
PR 23-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235634P.
PR 27-SEP-2000; 2000US-0235636P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-024060P.
PR 20-OCT-2000; 2000US-0241211P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241877P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.

Search completed: February 23, 2004, 19:50:07
Job time : 204.357 secs

```
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256199P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251855P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
DR WPI; 2001-465570/50.  
XX  
XX Isolated nucleic acid molecule encoding a reproductive system antigen is  
PT used in preventing, treating or ameliorating a medical condition.  
XX  
XX  
XX  
PS Disclosure; SEQ ID NO 6589; 1297bp + Sequence Listing; English.  
XX  
XX The present invention provides the protein and coding sequences of a  
CC number of human reproductive system related antigens. These can be used  
CC in the prevention and treatment of reproductive system disorders,  
CC including cancer. The present sequence is a genomic sequence encoding a  
CC protein of the invention  
XX  
XX Sequence 5049 BP; 1248 A; 1315 C; 1438 G; 1048 T; 0 U; 0 Other;  
SQ  
Query Match 9.2%; Score 46; DB 4; Length 5049;  
Best Local Similarity 57.7%; Pred. No. 0.027;  
Matches 82; Conservative 0; Mismatches 60; Indels 0; Gaps 0;  
QY 91 GGGCTCTGCTGCTCTCTCTCGAGAGGCGGTGAGTCTCCAGTATGTTGGCCAGC 150  
DB 1575 GGGGTGTTATCTCTGAGTAGGAGACAGAGTTGGGCTCTCAGATGCTGACACAGC 1516  
QY 151 GGTCTTCGACCTCTCTGAGCAAGCGCAAGAGCTGACGACGAGAGAGGGGGCTCA 210  
DB 1515 AGGGTCTTACCTCTGAGCCCAATGTATTAAGCACTGAGGCGGAGATTAAGGGCTTT 1456  
QY 211 CCTTTATCTCGGCGACCCAC 232  
DB 1455 AGAGTAGGCGCTGGAGGCTC 1434
```

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 102
LENGTH: 846
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-102

Query Match
Best Local Similarity 7.3%; Score 36.6; DB 4; Length 846;
Pred. No. 0.68;
Matches 60; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 318 GGACATGCGAGCGCGGGAAGAGTGCCTGCGGCGGCGGAGAGAGTCCAGAG 377
DB 727 GCGCGTCCGAGACCGGACGCTTCTGCTGCGGCGGCGGAGAGTGTGCGAGGTAC 786
QY 378 GGCACGTGAGCGCTGCTGCGCGGAGGCTTCGACATCCGC 416
DB 787 AGCTGACGCCCGAGCTTGCGCAAGGCTGGGCGCACCGGC 825

RESULT 3
US-09-252-991A-98
Sequence 98, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252.991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 98
LENGTH: 1038
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-98

Query Match
Best Local Similarity 7.3%; Score 36.6; DB 4; Length 1038;
Pred. No. 0.74;
Matches 60; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 318 GGACATGCGAGCGCGGGAAGAGTGCCTGCGGCGGCGGAGAGTCCAGAG 377
DB 737 GCGCGTCCGAGACCGGACGCTTCTGCTGCGGCGGCGGAGAGTGTGCGAGGTAC 796
QY 378 GGCACGTGAGCGCTGCTGCGCGGAGGCTTCGACATCCGC 416
DB 797 AGCTGACGCCCGAGCTTGCGCAAGGCTGGGCGCACCGGC 835

RESULT 4
US-09-252-991A-91/c
Sequence 91, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252.991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 91
LENGTH: 1398
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-91

TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-91

Query Match
Best Local Similarity 7.3%; Score 36.6; DB 4; Length 1398;
Pred. No. 0.83;
Matches 60; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 318 GGACATGCGAGCGCGGGAAGAGTGCCTGCGGCGGCGGAGAGTCCAGAG 377
DB 767 GCGCGTCCGAGACCGGACGCTTCTGCTGCGGCGGCGGAGAGTGTGCGAGGTAC 708
QY 378 GGCACGTGAGCGCTGCTGCGCGGAGGCTTCGACATCCGC 416
DB 707 AGCTGACGCCCGAGCTTGCGCAAGGCTGGGCGCACCGGC 669

RESULT 5
US-09-252-991A-84/c
Sequence 84, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252.991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 84
LENGTH: 2934
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-84

Query Match
Best Local Similarity 7.3%; Score 36.6; DB 4; Length 2934;
Pred. No. 1.1;
Matches 60; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 318 GGACATGCGAGCGCGGGAAGAGTGCCTGCGGCGGCGGAGAGTCCAGAG 377
DB 2055 GCGCGTCCGAGACCGGACGCTTCTGCTGCGGCGGCGGAGAGTGTGCGAGGTAC 1996
QY 378 GGCACGTGAGCGCTGCTGCGCGGAGGCTTCGACATCCGC 416
DB 1995 AGCTGACGCCCGAGCTTGCGCAAGGCTGGGCGCACCGGC 1957

RESULT 6
US-09-252-991A-3564/c
Sequence 3564, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252.991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 3564
LENGTH: 504
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3564

Query Match 7.2%; Score 36; DB 4; Length 504;
Best Local Similarity 51.9%; Pred. No. 0.81;
Matches 81; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 287 GGGGACCAATCTGAGTCTTCCCGCTTGGAATCGAGCGCGGAAAGCAATGCC 346
DB 193 GGGGAGCGCGCATGATCTCGACGAGCGCGCGCGCTTGAGGCTGGCGCC 134
QY 347 GCCAGGCTGGGCGCAGAGAGCTCCAGAAAGGCACTGAGCGCTGCGCGAGGCTC 406
DB 133 ACGGTGACGTCCCGGTGCTGCCCAAGCGCGCTGAGAGCGCTGTCGATCCATTTC 74
QY 407 GGACATCCGAGGACCAAGGAAAGTCTCTGCGGC 442
DB 73 GCCCATCCGAGGCGCTCGGGAAGATTCCGCGCGC 38

RESULT 7
US-09-252-991A-3539
Sequence 3539, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 3539
LENGTH: 1002
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3539

Query Match 7.2%; Score 36; DB 4; Length 1002;
Best Local Similarity 51.9%; Pred. No. 1.1;
Matches 81; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 287 GGGGACCAATCTGAGTCTTCCCGCTTGGAATCGAGCGCGGAAAGCAATGCC 346
DB 281 GGGGAGCGCGCATGATCTCGACGAGCGCGCGCGCTTGAGGCTGGCGCC 340
QY 347 GCCAGGCTGGGCGCAGAGAGCTCCAGAAAGGCACTGAGCGCTGCGCGAGGCTC 406
DB 341 ACGGTGACGTCCCGGTGCTGCCCAAGCGCGCTGAGAGCGCTGTCGATCCATTTC 400
QY 407 GGACATCCGAGGACCAAGGAAAGTCTCTGCGGC 442
DB 401 GCCCATCCGAGGCGCTCGGGAAGATTCCGCGCGC 436

RESULT 8
US-09-252-991A-3548/c
Sequence 3548, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 3548
LENGTH: 1644
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3548

Query Match 7.2%; Score 36; DB 4; Length 1644;
Best Local Similarity 51.9%; Pred. No. 1.3;
Matches 81; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 287 GGGGACCAATCTGAGTCTTCCCGCTTGGAATCGAGCGCGGAAAGCAATGCC 346
DB 816 GGGGAGCGCGCATGATCTCGACGAGCGCGCGCGCTTGAGGCTGGCGCC 757
QY 347 GCCAGGCTGGGCGCAGAGAGCTCCAGAAAGGCACTGAGCGCTGCGCGAGGCTC 406
DB 756 ACGGTGACGTCCCGGTGCTGCCCAAGCGCGCTGAGAGCGCTGTCGATCCATTTC 697
QY 407 GGACATCCGAGGACCAAGGAAAGTCTCTGCGGC 442
DB 696 GCCCATCCGAGGCGCTCGGGAAGATTCCGCGCGC 661

RESULT 9
US-09-252-991A-3523
Sequence 3523, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 3523
LENGTH: 2832
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3523

Query Match 7.2%; Score 36; DB 4; Length 2832;
Best Local Similarity 51.9%; Pred. No. 1.6;
Matches 81; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 287 GGGGACCAATCTGAGTCTTCCCGCTTGGAATCGAGCGCGGAAAGCAATGCC 346
DB 1054 GGGGAGCGCGCATGATCTCGACGAGCGCGCGCGCTTGAGGCTGGCGCC 1113
QY 347 GCCAGGCTGGGCGCAGAGAGCTCCAGAAAGGCACTGAGCGCTGCGCGAGGCTC 406
DB 1114 ACGGTGACGTCCCGGTGCTGCCCAAGCGCGCTGAGAGCGCTGTCGATCCATTTC 1173
QY 407 GGACATCCGAGGACCAAGGAAAGTCTCTGCGGC 442
DB 1174 GCCCATCCGAGGCGCTCGGGAAGATTCCGCGCGC 1209

RESULT 10
US-09-252-991A-11606
Sequence 11606, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18

Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 15239
LENGTH: 810
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15239

Query Match 7.0%; Score 35; DB 4; Length 810;
Best Local Similarity 54.2%; Pred. No. 1.9;
Matches 71; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 318 GGGACATCGGAGCGGAGGAGCGAGGCTGAGGCTGAGGAGAGAGCTCCAGAG 377
DB 176 GGGATTGCGGAGCGGAGCTGAGGCGGCGGACCTGCTGAGGCTCGGCGGCG 235
QY 378 GGCACGTGAGCGCTGCGCGGAGGCTCGGACATCGGAGGAGGAGGAGAAAGTCTCT 437
DB 236 GCGGTTGCGCGGCTGAGCGGAGGCTCGGAGGCTCGGAGGCTCGGAGGCTCGG 295
QY 438 GGGGCGATCTG 448
DB 296 GTGGCGAATCG 306

RESULT 15

US-09-252-991A-15234
Sequence 15234; Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 15234
LENGTH: 915
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15234

Query Match 7.0%; Score 35; DB 4; Length 915;
Best Local Similarity 54.2%; Pred. No. 2;
Matches 71; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 318 GGGACATCGGAGCGGAGGAGCGAGGCTGAGGCTGAGGAGAGAGCTCCAGAG 377
DB 54 GGGATTGCGGAGCGGAGCTGAGGCGGCGGACCTGCTGAGGCTCGGCGGCG 113
QY 378 GGCACGTGAGCGCTGCGCGGAGGCTCGGACATCGGAGGAGGAGGAGAAAGTCTCT 437
DB 114 GCGGTTGCGCGGCTGAGCGGAGGCTCGGAGGCTCGGAGGCTCGGAGGCTCGG 173
QY 438 GGGGCGATCTG 448
DB 174 GTGGCGAATCG 184


```

QY 181 AAGAGCTGACGACAGAGAGAGAGGGGCTCACCTCTTATCCTCGGCAACCACTGCACAA 240
Db 1143 AAGAGCTGACGACAGAGAGAGGGGCTCACCTCTTATCCTCGGCAACCACTGCACAA 1202
QY 241 CAGGCGCGCTCTCCGACACTTAAATGTATCAACCACTAACCTGTATGAGGGGAGCCCAATCTG 300
Db 1203 CAGGCGCGCTCTCCGACACTTAAATGTATCAACCACTAACCTGTATGAGGGGAGCCCAATCTG 1262
QY 301 GACTCCTTCCCGCGCTTGGAGCATGCAAGCGCGGAGAACATGCGCCGCAAGCCTTGAGCC 360
Db 1263 GACTCCTTCCCGCGCTTGGAGCATGCAAGCGCGGAGAACATGCGCCGCAAGCCTTGAGCC 1322
QY 361 AAGAGAGCTCCAGAGAGAGGACATGAGCGCTGTGCGGAGAGCCTCGGACATCGCAGGC 420
Db 1323 AAGAGAGCTCCAGAGAGAGGACATGAGCGCTGTGCGGAGAGCCTCGGACATCGCAGGC 1382
QY 421 ACCAGGGAAGTCTCCTGGGCGCATCTGTAAATAAACCCTTTTCTTTTGTGTTTAA 480
Db 1383 ACCAGGGAAGTCTCCTGGGCGCATCTGTAAATAAACCCTTTTCTTTTGTGTTTAA 1442
QY 481 AAAAAATTAAAA 491
Db 1443 AAAAAATTAAAA 1453

```

RESULT 2

```

US-10-187-657-2
; Sequence 2, Application US/10187657
; Publication No. US20030068311A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy K.W.
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Azimzai, Yalda
; TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
; FILE REFERENCE: PV-0009 CIP
; CURRENT APPLICATION NUMBER: US/10/187,657
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: PCT/US00/07817
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/139,565
; PRIOR FILING DATE: 1999-06-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program
; SEQ ID NO 2
; LENGTH: 1421
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030068311A1 4901066GB1
US-10-187-657-2

```

```

Query Match          91.0%; Score 453; DB 14; Length 1421;
Best Local Similarity 99.6%; Pred. No. 1.1e-126;
Matches 475; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

```

```

QY 1 CTTAGCGTGCCTGCTGCGCGCTTCCGCTTAGGCTCTTCCGCGTCAACATCAGTACG 60
Db 947 CTTAGCGTGCCTGCTGCGCGCTTCCGCTTAGGCTCTTCCGCGTCAACATCAGTACG 1005
QY 61 AAGCGCGCTTCTGGGATCAAGCTGGAACCGGCGCTGTGCTCTTCTGAGAGGGGCC 120
Db 1006 -GGCGCGCTTCTGGGATCAAGCTGGAACCGGCGCTGTGCTCTTCTGAGAGGGGCC 1064
QY 121 GTGTGAGTCTCCAGTATGTTGGCCCAAGCGCTCTTGGACACCTTCTGAGCAAAAGGCC 180
Db 1065 GTGTGAGTCTCCAGTATGTTGGCCCAAGCGCTCTTGGACACCTTCTGAGCAAAAGGCC 1124
QY 181 AAGGACTGACGACAGAGAGAGGGGCTCACTCTTATCCTCGGAGAACCACTGCACAA 240
Db 1125 AAGGACTGACGACAGAGAGAGGGGCTCACTCTTATCCTCGGAGAACCACTGCACAA 1184
QY 241 CAGGCGCTCTCCAGACTTAAATGTATCAACCACTAACCTGTGAGGGGAGCCCAATCTG 300

```

```

Db 1185 CAGGCGCGCTCTCCAGACTTAAATGTATCAACCACTAACCTGTGAGGGGAGCCCAATCTG 1244
QY 301 GACTCCTTCCCGCGCTTGGAGCATGCAAGCGCGGAGAACATGCGCCGCAAGCCTTGAGCC 360
Db 1245 GACTCCTTCCCGCGCTTGGAGCATGCAAGCGCGGAGAACATGCGCCGCAAGCCTTGAGCC 1304
QY 361 AAGAGAGCTCCAGAGAGAGGACATGAGCGCTGTGCGGAGAGCCTCGGACATCGCAGGC 420
Db 1305 AAGAGAGCTCCAGAGAGAGGACATGAGCGCTGTGCGGAGAGCCTCGGACATCGCAGGC 1364
QY 421 ACCAGGGAAGTCTCCTGGGCGCATCTGTAAATAAACCCTTTTCTTTTGTGTTT 477
Db 1365 ACCAGGGAAGTCTCCTGGGCGCATCTGTAAATAAACCCTTTTCTTTTGTGTTT 1421

```

RESULT 3

```

US-10-187-657-4
; Sequence 4, Application US/10187657
; Publication No. US20030068311A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy K.W.
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Azimzai, Yalda
; TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
; FILE REFERENCE: PV-0009 CIP
; CURRENT APPLICATION NUMBER: US/10/187,657
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: PCT/US00/07817
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/139,565
; PRIOR FILING DATE: 1999-06-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program
; SEQ ID NO 4
; LENGTH: 1420
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030068311A1 3221661CA2
US-10-187-657-4

```

```

Query Match          89.3%; Score 447.8; DB 14; Length 1420;
Best Local Similarity 99.2%; Pred. No. 4.2e-125;
Matches 471; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

```

```

QY 1 CTTAGCGTGCCTGCTGCGCGCTTCCGCTTAGGCTCTTCCGCGTCAACATCAGTACG 60
Db 947 CTTAGCGTGCCTGCTGCGCGCTTCCGCTTAGGCTCTTCCGCGTCAACATCAGTACG 1005
QY 61 AAGCGCGCTTCTGGGATCAAGCTGGAACCGGCGCTGTGCTCTTCTGAGAGGGGCC 120
Db 1006 -GGCGCGCTTCTGGGATCAAGCTGGAACCGGCGCTGTGCTCTTCTGAGAGGGGCC 1064
QY 121 GTGTGAGTCTCCAGTATGTTGGCCCAAGCGCTCTTGGACACCTTCTGAGCAAAAGGCC 180
Db 1065 GTGTGAGTCTCCAGTATGTTGGCCCAAGCGCTCTTGGACACCTTCTGAGCAAAAGGCC 1124
QY 181 AAGGACTGACGACAGAGAGAGGGGCTCACTCTTATCCTCGGAGAACCACTGCACAA 240
Db 1125 AAGGACTGACGACAGAGAGAGGGGCTCACTCTTATCCTCGGAGAACCACTGCACAA 1184
QY 241 CAGGCGCTCTCCAGACTTAAATGTATCAACCACTAACCTGTGAGGGGAGCCCAATCTG 300
Db 1185 CAGGCGCTCTCCAGACTTAAATGTATCAACCACTAACCTGTGAGGGGAGCCCAATCTG 1244
QY 301 GACTCCTTCCCGCGCTTGGAGCATGCAAGCGCGGAGAACATGCGCCGCAAGCCTTGAGCC 360
Db 1245 GACTCCTTCCCGCGCTTGGAGCATGCAAGCGCGGAGAACATGCGCCGCAAGCCTTGAGCC 1304
QY 361 AAGAGAGCTCCAGAGAGAGGACATGAGCGCTGTGCGGAGAGCCTCGGACATCGCAGGC 420

```


RESULT 7
US-10-027-632-91970
Sequence 91970. Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
POLYMORPHISMS in the Human Genome

```

RESULT 8
US-10-123-155-10/c
Sequence 10, Application US/10123155
Publication No. US20030068794A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerlitsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SCDATED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3330R1C30
CURRENT APPLICATION NUMBER: US/10/123,155
CURRENT FILING DATE: 2002-04-15
Prior Application removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 10
LENGTH: 594

```



```

? ORGANISM: Homo Sapien
US-10-146-731-10
Query Match      10.8%; Score 49.8; DB 14; Length 594;
Best Local Similarity 7.5%; Pred. No. 9.7e-05;
Matches 36; Conservative 183; Mismatches 260; Indels 0; Gaps 0;

QY   6 GGCTGCCGCTCGCCGGTCGCCGTGCAGTCTCCTCGCGGGTGATCAACCATGACGAAGGCG 65
DB   : :::: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DY   507 SYSYS.S.:SMSYSISSSDDI.CTCCTYRHGDSTSYSITY.CRCCTYT.SYSRDCXY 448
QY   66 CGCTTTCTGAGTCACGCTGCGCACCGGCGTCTGCTCTTCTTCGAGAAGGCGCGT 125
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DY   447 SCCGSDCYGSYSRSRYSYSYSWSWISYSTDYSYNKCCCYBSISSSYSSSAY 388
QY   126 GAGTCTCCAGTAGTGTGGGCCAAGCGTCTTCGCACCCTTTGACACCAAAGGCCACA 185
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DY   387 STSSSSSSSSSYITNTYC.T.CC...T.MCAABGSTTTTTTTTT.HSCC.SA.. 328
QY   186 CTGACGACGAGGAGAGAGGAGGCTCACTCTATCTCGGAGAACCCATGACAGAGCG 245
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DY   327 A.M.YC.A.SYSYSYS.SSS.S.SYMR.HRA.SHYTS.S.WCY YM.Y.YY.XY 268
QY   246 GCCTCTCCCAGACTTAATAATGATACCACTAACTGTGAGGAGGACCAANTGCACTC 305
DB   267 SYCCSRKTM...TMTDM.T.T.T.MMY.RYB.HCHTKCAAT.NN.HTB...N.HHBH 208
QY   306 CTTCCCCCTTTGGGACATGACAGGCGGAGAGAGCGCCGACGCGTGGGACAGAG 365
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DY   207 .BH.H.B.H.HNS..TTS....M.TTM.B.TBAS.H.RAFC.STYM.T.MM.H.S. 148
QY   366 AGCTCCAGAAAGGACATGACCGCTGTCGGCGGAGAGCCCTGGACATCGCAGGACGAG 425
DB   147 Y..CCR.H.R..R.G...S.CT HN.B.CYRW.RUNGY.HS...S.AA...ARH.D 88
QY   426 GGAAGTCTCCTGGGGCGCATGTGAATAAACCTTTTCTTTCTTTTAAAAAAA 464
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DY   87 H.DTYBH.YH.KNNAH.C.W.KCNB.S.IT.TBTSH.DEMGTICDB...T.HMM.M.RR 29

RESULT 10
US-10-140-472-10/C
? Sequence 10, Application US/10140472
? Publication No. US2003013888BA1
GENERAL INFORMATION:
APPLICANT: Baker,Kevin P.
APPLICANT: Beresini,Mareeen
APPLICANT: DeForge,Laura
APPLICANT: Desnoyers,Luc
APPLICANT: Filvaroff,Ellen
APPLICANT: Gao,Wel-Qiang
APPLICANT: Gerritsen,Mary E.
APPLICANT: Goddard,Audrey J.
APPLICANT: Godowski,Paul J.
APPLICANT: Gurney,Austin L.
APPLICANT: Sherwood,Steven
APPLICANT: Smith,Victoria
APPLICANT: Stewart,Timothy A.
APPLICANT: Tumas,Daniel
APPLICANT: Watanabe,Colin K
APPLICANT: Wood,William
APPLICANT: Zhang,Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3350RIC168
CURRENT APPLICATION NUMBER: US/10/140.472
Prior Filing DATE: 2002-05-06
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 10
LENGTH: 594
TYPE: PRT
ORGANISM: Homo Sapien
```

US-10-140-472-10

Query Match 10.0%; Score 49.8; DB 14; Length 594;
 Best Local Similarity 7.5%; Pred. No. 9.7e-05;
 Matches 36; Conservative 183; Mismatches 260; Indels 0; Gaps 0

[illegible]

RESULT 11

US-10-141-761-10/c
; Sequence 10, Application US/10141761

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
 APPLICANT: Beresini, Maureen
 APPLICANT: DeForge, Laura
 APPLICANT: Demoyers, Luc
 APPLICANT: Flivaroff, Ellen
 APPLICANT: Gao, Wei-Qiang
 APPLICANT: Gertlesen, Mary E.
 APPLICANT: Goddard, Audrey
 APPLICANT: Goddowski, Paul J.
 APPLICANT: Gurney, Austin L.
 APPLICANT: Sherwood, Steven
 APPLICANT: Smith, Victoria A.
 APPLICANT: Stewart, Timothy A.
 APPLICANT: Tamas, Daniel
 APPLICANT: Watanabe, Colin K
 APPLICANT: Wood, William
 APPLICANT: Zhang, Zemin
 TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 TITLE OF INVENTION: ACIDS ENCODING THE SAME
 FILE REFERENCE: P33ORC1A198
 CURRENT APPLICATION NUMBER: US/10/141,761
 CURRENT FILING DATE: 2002-05-08
 Prior Application removed - See Palm or File Wrapper
 NUMBER OF SEQ ID NOS: 550

Query Match: 10.0%; Score 49.8; DB 14; Length 594;
Best Local Similarity 7.5%; Pred. No. 9.7e-05;
Matches 36; Conservative 183; Mismatches 260; Indels 0; Gaps 0;

```

QY      GCGTGGCGGCTCCCGCGCTCCCGCTGAGCGCTCCCGGGCTCAGCATTGACAGTGCAGGCG 65
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      507 SYIS.S.SWMSYSISSDDY.CYCTCYRHCSISYSIY.CRCCTYT.SISRDCHY 448
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      66 CGCCTTCTGGGTACGCGCTGGCAACCGCGCTGCTGCGCTTCTCTCGAGAGGCGCTGCT 125
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      447 SCCGSDPYCYSYSARYSYSYSYSWMSYSDYDYSYRCCCTCYSSYSISSYSSAY 388
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      126 GAGTCTTCAGTATGTTGGGCCGAGGCTCTTCGCACTTGTGACCCAAAGCGCCAAAGA 185
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      387 STSSSSSSSSYYIYSTNYC.T.CC...T.MCAABSTTTTTTTTTT..HSQC.SA.. 328
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      186 CTGCACCCGAGAGAGAGGCGGCTCCTCTTATCTTCGCGCAGCCATGCAAGCAGCG 245
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      327 A.M..YC.A.SYSIS.SS.S.SYM.HBA.SHYIRS..S.MCY.YM.Y..Y.Y 268
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      246 CGCTCTCCAGACTTAATGTATCATCACCACTAACCTGTGAGGGGAGCCCATCTGACTC 305
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      267 SYYSCKRTM...TMTDM.T.T..MHMY.KTB.HCHTKCRAT.MN.HTB..N..HBHB 208
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      306 CTTCGCCGCGCTTGGGACATCGCAGGCGGGAGAGACGCGCCGACGGCGTGGGCGAGAG 365
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      207 .BH.H..B.H.HSNs...TTS....M.TTM.B.TBASH.RARC.STMM.T.MMM.H.S. 148
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      366 AGCTCAGGAAGAGGACATGAGCGCTGCGGCGAGGCTTCGACATCGCAGGCAACAG 425
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      147 Y..CCR..H.R..R.G...S.CT.HN.B.CVRV.RNGMY.HS...S.AA...ARH..D 88
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      426 GGAAGATCTCCCTGGGGCGATCTGTAAATAAACCCTTTTCTTTGTTTAAAAAAA 484
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      87 H..DTYSH.YH.KNNAH.C.W.KKNB..S.TT.TBTSH.DBMGTVCDB...T.HMM.M.RR 29
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 12

US-10-142-885-10/c
; Sequence 10, Application US/10142885

GENERAL INFORMATION:

```

1  APPLICANT: Baker, Kevin P.
2  APPLICANT: Beresini, Naureen
3  APPLICANT: Deforge, Laura
4  APPLICANT: Desnoyers, Luc
5  APPLICANT: Filvaroff, Ellen
6  APPLICANT: Gao, Wei-Qiang
7  APPLICANT: Gerlitsen, Mary E.
8  APPLICANT: Goddard, Audrey
9  APPLICANT: Godowski, Paul J.
10 APPLICANT: Gurney, Austin L.
11 APPLICANT: Sherwood, Steven
12 APPLICANT: Smith, Victoria
13 APPLICANT: Stewart, Timothy A.
14 APPLICANT: Tamas, Daniel
15 APPLICANT: Watanabe, Colin K.
16 APPLICANT: Wood, William
17 APPLICANT: Zhang, Zhenli
18
19 TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
20 ACIDS ENCODING THE SAME
21 FILE REFERENCE: P333081C248
22 CURRENT APPLICATION NUMBER: US/10/142,885
23 CURRENT FILING DATE: 2002-05-10
24 Prior Application removed - See File Wrapper or Palm
25 NUMBER OF SEQ ID NOS: 550

```

Query Match 10.0%; Score 49.8; DB 14; Length 594;
Best Local Similarity 7.5%; Pred. No. 9,7e-05;
Matches 36; Conservative 183; Mismatches 260; Indels 0; Gaps 0;

```
QY 6 GCGTCCGCTGCGCCGCTCCGCTAGGCTCTCCGCGCTACCACTGATGACGCGC 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 507 SYSYS.S.SWSYSYSSSDDY.CYCCYRHHCDSSYSYXY.CRCCTYT.SYSRDCHY 448
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 66 CGCCTTCGAGTCAAGCGGACCGGCGCTGCTCTCTCTCGGAGGCGCTGCT 125
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 447 SCCSDYCYYSRYSRYSYSYSWSYSYTDYCYRCCCTYSISSYSISSSNAY 388
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 126 GAGTCTCCAGTATGTTGCGCCAGCGCTCTTGCGACCCCTTGACCAAGCGCAAGA 185
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 387 STSSSSSSSSSYTYSTNYC.T.CC...T.MCABGCTTTT...HSCC.SA.. 328
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 186 CTGACGCGAGAGAGAGGGGCTCACCCTTATCTCTGCGGACCCACTGACAGAGAGC 245
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 327 A.M.YC.A.SYSYSYS.SSS.S.SYWR.HRA.SHYTTS.S.MTCY.YM.Y.YY.YY 268
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 246 CGCTCTCCAGACTTAAATGTATCACCACCTAAGCTGTGAGGGGACCCCAATCTGACTC 305
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 267 SYCSRKTM...TMTDM.T.T.MMY.KYB.HCHTKRAAT.MN.HTB..N.HHB 208
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 306 CTTCGCCGCTTGAGACATGCGAGCGCGGAGAGAGTCCCGCAGGCTGAGCCAGAG 365
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 207 .BH.H..B.H.HNS...TTS...M.TTM.B.TBASH.RARC.STMM.T.MMM.H.S. 148
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 366 AGCTCCAGAGAGGCACTGAGCGCTGCTGCGGAGAGCTCGAGACATCCGAGAGCAG 425
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 147 Y.CCR..H.R.R.G...S.CT.HN.B.CYR.RNMY.HS...S.AA...AH..D 88
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 426 GGAAGTCTCTGCGGCGCATGTGTAATTAACCTTTTCTTTGTTTAAAAAA 484
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 87 H.DTYBH.YH.XNNAH.C.W.KKMB.S.TT.TBTSB.DMGTYCDB...T.HMM.M.RR 29
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 13

US-10-158-790-10/c
Sequence 10, Application US/10158790
Publication No. US20030180879A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: Desnoyers, Luc
APPLICANT: Flivaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P330R1C448
CURRENT APPLICATION NUMBER: US/10/158,790
CURRENT FILING DATE: 2002-05-30
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 10
LENGTH: 594
TYPE: PRT
ORGANISM: Homo Sapien
US-10-158-790-10

Query Match 10.0%; Score 49.8; DB 14; Length 594;

Best Local Similarity 7.5%; Pred. No. 9,7e-05;
Matches 36; Conservative 183; Mismatches 260; Indels 0; Gaps 0;

```
QY 6 GCGTCCGCTGCGCCGCTCCGCTAGGCTCTCCGCGCTACCACTGATGACGCGC 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 507 SYSYS.S.SWSYSYSSSDDY.CYCCYRHHCDSSYSYXY.CRCCTYT.SYSRDCHY 448
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 66 CGCCTTCGAGTCAAGCGGACCGGCGCTGCTCTCTCTCGGAGGCGCTGCT 125
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 447 SCCSDYCYYSRYSRYSYSYSWSYSYTDYCYRCCCTYSISSYSISSSNAY 388
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 126 GAGTCTCCAGTATGTTGCGCCAGCGCTCTTGCGACCCCTTGACCAAGCGCAAGA 185
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 387 STSSSSSSSSSYTYSTNYC.T.CC...T.MCABGCTTTT...HSCC.SA.. 328
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 186 CTGACGCGAGAGAGAGGGGCTCACCCTTATCTCTGCGGACCCACTGACAGAGAGC 245
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 327 A.M.YC.A.SYSYSYS.SSS.S.SYWR.HRA.SHYTTS.S.MTCY.YM.Y.YY.YY 268
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 246 CGCTCTCCAGACTTAAATGTATCACCACCTAAGCTGTGAGGGGACCCCAATCTGACTC 305
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 267 SYCSRKTM...TMTDM.T.T.MMY.KYB.HCHTKRAAT.MN.HTB..N.HHB 208
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 306 CTTCGCCGCTTGAGACATGCGAGCGCGGAGAGAGTCCCGCAGGCTGAGCCAGAG 365
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 207 .BH.H..B.H.HNS...TTS...M.TTM.B.TBASH.RARC.STMM.T.MMM.H.S. 148
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 366 AGCTCCAGAGAGGCACTGAGCGCTGCTGCGGAGAGCTCGAGACATCCGAGAGCAG 425
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 147 Y.CCR..H.R.R.G...S.CT.HN.B.CYR.RNMY.HS...S.AA...AH..D 88
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 426 GGAAGTCTCTGCGGCGCATGTGTAATTAACCTTTTCTTTGTTTAAAAAA 484
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 87 H.DTYBH.YH.XNNAH.C.W.KKMB.S.TT.TBTSB.DMGTYCDB...T.HMM.M.RR 29
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 14

US-10-137-871-10/c
Sequence 10, Application US/10137871
Publication No. US20030207350A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: Desnoyers, Luc
APPLICANT: Flivaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P330R1C153
CURRENT APPLICATION NUMBER: US/10/137,871
CURRENT FILING DATE: 2002-05-03
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 10
LENGTH: 594
TYPE: PRT
ORGANISM: Homo Sapien
US-10-137-871-10

Query Match 10.0%; Score 49.8; DB 15; Length 594;
Best Local Similarity 7.5%; Pred. No. 9,7e-05;

Matches 36; Conservative 183; Mismatches 260; Indels 0; Gaps 0;

```

QY 6 GGGTGGCGCTGTCGCCGCTCCGCTAGGCTCTCCGCTCAGCACTGATGACGAGCC 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 507 SYSYS.S.SMSYSYSSSDY.CYCCYRHCHSDYSYSYY.CRCYCT.SYSRDYCHY 448
QY 66 CGCTTCTGGGTACAGCTGGCAAGCGGCTGCTGCTCTCTCTCGAGAGGCGCTGT 125
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 447 SCCGSDYCYCYSYSYSYSYSYSYSYTDYCSYRCCYYSYSSYSYSSSAY 388
QY 126 GAGTCTCCAGTATGTTGGCCGAGCTCTTCCGACCCCTTCCGACCAAGCGCAGGA 185
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 387 STSSSSSSSYTYSNYC.T.CC...T.MCAABCTTTTTTTTTT.HSCC.SA. 328
QY 186 CTGACGCCAGAGAGAGAGGAGGCTCCTTATCTCGGAGCCACTGACAGAGCC 245
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 327 A.M..YC.A.SYSYSYS.SSS.S.SYMR.HRA.SHYTYS..S.MYCY.YM.Y.Y.Y 268
QY 246 CGCTCTCCAGACTTAATGATATACACACTAAGCTGAGGGGAGCCCAATCTGACTC 305
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 267 SYCSRKTM...TMTDM.T.T.MMYT.KTB.HCHTKRAAT.MN.HTB...N.HBHB 208
QY 306 CTTCGCCGCTTGGACATGCGAGCGGAGAGCACTGCCGCGAGCTTGGCCAGAG 365
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 207 .BH.H..B.H.HSNS...TTS...M.TTM.B.TBASH.RARC.STMM.T.MMM.H.S. 148
QY 366 AGCTCCAGAGAGGAGGACTGAGCGCTGCTGCGCGAGGCTCGACATCCGAGCAGCAG 425
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 147 Y..CCR..H.R..R.G...S.CT.HN.B.CYRY.RNGMY.HS...S.AA...ARH..D 88
QY 426 GGAAGTCTCTGGGCGAGTCTGTAATAACCTTTTCTTTTGTGTTTTTAAAAA 484
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 87 H.DTYBH.YH.KNNAH.C.W.KXNB.S.TT.TBISH.DMGTYCDB...T.HMM.M.RR 29

```

RESULT 15

```

US-10-140-923-10/c
; Sequence 10, Application US/10140923
; Publication No. US20030207355A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerltzen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Goddard, Paul J.
; APPLICANT: Guirney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C188
; CURRENT APPLICATION NUMBER: US/10/140,923
; PRIOR APPLICATION REMOVED - 2002-05-07
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 10
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-923-10

```

Query Match 10.0%, Score 49.8, DB 15, Length 594;
 Best Local Similarity 7.5%, Ptd. No. 9.7e-05;
 Matches 36; Conservative 183; Mismatches 260; Indels 0; Gaps 0;

Search completed: February 23, 2004, 23:36:21
 Job time: 202.084 secs

```

QY 6 GGGTGGCGCTGTCGCCGCTCCGCTAGGCTCTCCGCTCAGCACTGATGACGAGCC 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 507 SYSYS.S.SMSYSYSSSDY.CYCCYRHCHSDYSYSYY.CRCYCT.SYSRDYCHY 448
QY 66 CGCTTCTGGGTACAGCTGGCAAGCGGCTGCTGCTCTCTCTCGAGAGGCGCTGT 125
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 447 SCCGSDYCYCYSYSYSYSYSYSYTDYCSYRCCYYSYSSYSYSSSAY 388
QY 126 GAGTCTCCAGTATGTTGGCCGAGCTCTTCCGACCCCTTCCGACCAAGCGCAGGA 185
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 387 STSSSSSSSYTYSNYC.T.CC...T.MCAABCTTTTTTTTTT.HSCC.SA. 328
QY 186 CTGACGCCAGAGAGAGAGGAGGCTCCTTATCTCGGAGCCACTGACAGAGCC 245
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 327 A.M..YC.A.SYSYSYS.SSS.S.SYMR.HRA.SHYTYS..S.MYCY.YM.Y.Y.Y 268
QY 246 CGCTCTCCAGACTTAATGATATACACACTAAGCTGAGGGGAGCCCAATCTGACTC 305
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 267 SYCSRKTM...TMTDM.T.T.MMYT.KTB.HCHTKRAAT.MN.HTB...N.HBHB 208
QY 306 CTTCGCCGCTTGGACATGCGAGCGGAGAGCACTGCCGCGAGCTTGGCCAGAG 365
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 207 .BH.H..B.H.HSNS...TTS...M.TTM.B.TBASH.RARC.STMM.T.MMM.H.S. 148
QY 366 AGCTCCAGAGAGGAGGACTGAGCGCTGCTGCGCGAGGCTCGACATCCGAGCAGCAG 425
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 147 Y..CCR..H.R..R.G...S.CT.HN.B.CYRY.RNGMY.HS...S.AA...ARH..D 88
QY 426 GGAAGTCTCTGGGCGAGTCTGTAATAACCTTTTCTTTTGTGTTTTTAAAAA 484
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 87 H.DTYBH.YH.KNNAH.C.W.KXNB.S.TT.TBISH.DMGTYCDB...T.HMM.M.RR 29

```

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	457.9	91.9	486	9	A1650560	w68bD08.x
C 2	456	91.6	515	9	A1828629	wR35e04.x
C 3	453	91.0	510	9	A1304327	q057U06.x
C 4	446.6	89.7	537	9	A1821606	nK08a11.x

C	5	433	87.3	446	A1984141
C	6	431.4	86.6	466	AA573825
C	7	423	84.9	434	A1991272
C	8	414	83.1	414	AA593860
C	9	409.4	82.2	411	AA583660
C	10	405	81.3	405	A1987893
C	11	400.2	80.4	405	A1274929
C	12	394.4	79.2	538	A1346155
C	13	387.4	77.8	573	A1660493
C	14	381	76.5	381	CB854121
C	15	378.2	75.9	737	A1281211
C	16	335.8	67.4	12	CB305339
C	17	335	67.3	348	BM977700
C	18	326.4	65.5	687	AM050605
C	19	316.8	63.6	324	CB850544
C	20	299.2	60.1	315	BM987789
C	21	296	59.4	296	AA615335
C	22	277	53.1	277	A1821178
C	23	264.2	53.1	505	A1732165
C	24	260	52.2	260	A1791844
C	25	252	50.6	252	AM009962
C	26	246.2	49.4	274	A1262415
C	27	238.2	48.4	274	AA465031
C	28	227	48.0	593	AA696031
C	29	218.4	38.4	539	BM973444
C	30	172.2	34.6	539	A1924216
C	31	158.2	31.6	713	A1924216
C	32	144.4	29.0	11	CEB117450
C	33	142	28.5	151	AK018459
C	34	138	27.7	408	AK008816
C	35	135.4	27.2	602	A1695625
C	36	94.4	19.0	378	B6958110
C	37	89	17.9	119	BQ917041
C	38	79.4	15.9	538	BM912889
C	39	54.8	11.0	925	AX5121281
C	40	52.8	10.6	982	CNS00091P
C	41	51.4	10.3	925	CNS45111
C	42	46	9.2	479	CNS00091P
C	43	45.2	9.1	1201	H83064
C	44	43.6	8.8	1000	AX357732
C	45	42.6	8.6	1201	BX407619
C	46	42.6	8.6	1269	BX350624
C	47	42.6	8.6	1269	BX450624
C	48	42.6	8.6	1269	BX506024
C	49	42.6	8.6	1269	BX606024
C	50	42.6	8.6	1269	BX706024
C	51	42.6	8.6	1269	BX806024
C	52	42.6	8.6	1269	BX906024
C	53	42.6	8.6	1269	BX006024
C	54	42.6	8.6	1269	BX106024
C	55	42.6	8.6	1269	BX206024
C	56	42.6	8.6	1269	BX306024
C	57	42.6	8.6	1269	BX406024
C	58	42.6	8.6	1269	BX506024
C	59	42.6	8.6	1269	BX606024
C	60	42.6	8.6	1269	BX706024
C	61	42.6	8.6	1269	BX806024
C	62	42.6	8.6	1269	BX906024
C	63	42.6	8.6	1269	BX006024
C	64	42.6	8.6	1269	BX106024
C	65	42.6	8.6	1269	BX206024
C	66	42.6	8.6	1269	BX306024
C	67	42.6	8.6	1269	BX406024
C	68	42.6	8.6	1269	BX506024
C	69	42.6	8.6	1269	BX606024
C	70	42.6	8.6	1269	BX706024
C	71	42.6	8.6	1269	BX806024
C	72	42.6	8.6	1269	BX906024
C	73	42.6	8.6	1269	BX006024
C	74	42.6	8.6	1269	BX106024
C	75	42.6	8.6	1269	BX206024
C	76	42.6	8.6	1269	BX306024
C	77	42.6	8.6	1269	BX406024
C	78	42.6	8.6	1269	BX506024
C	79	42.6	8.6	1269	BX606024
C	80	42.6	8.6	1269	BX706024

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION
AI660560/c	AI660560	486 bp mRNA linear EST 18-DEC-1999 weeb08.x1 Soares Dieckgrafe colon NHCD Homo sapiens cDNA clone IMAGE:2346231 3' similar to contains TAR1.t3 TAR1 repetitive element ;, mRNA sequence.

ACCESSION	AI660560
VERSION	AI660560.1
KEYWORDS	GI:4764130
SOURCE	EST.
	Homo sapiens (human)

Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 486)

TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP)
JOURNAL	Tumor Gene Index
COMMENT	Unpublished (1997)
	Contact: Robert Strausberg, Ph.D.

Email: CGABDS-remail.nlm.nih.gov
 This clone is available royalty-free through LNLN; contact the
 IMACS Consortium (infoimage@liln.gov) for further information.
 Insert Length: 143 Std Error: 0.00
 Seq primer: -400P from Gibco
 High quality sequence stop: 447.
 Location/Qualifiers
 1..486
 source

QY 421 ACCAGGAAAGTCTCTGCGGCGATCTGTAATTAACCTTTTCTTTTGTGTTTAA 480
 Db 60 ACCAGGAAAGTCTCTGCGGCGATCTGTAATTAACCTTTTCTTTTGTGTTTAA 1
 RESULT 3
 AI304327/c 510 bp mRNA linear EST 01-FEB-1999
 LOCUS g057h06.x1 NCI CGAP C08 Homo sapiens cDNA clone IMAGE:1912667 3'
 DEFINITION similar to contains TAR1.t3 TAR1 repetitive element ;, mRNA
 sequence.
 ACCESSION AI304327 GI:3988016
 VERSION AI304327.1
 KEYWORDS EST;
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 510)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 Tumor Gene Index
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Unpublished (1997)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
 www-bio.llnl.gov/bdrp/image/image.html
 Insert Length: 1487 Std Error: 0.00
 Seq primer: -400P from Gibco
 High quality sequence stop: 476.
 Location/Qualifiers
 1..510
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1912667"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B"
 /clone_lib="NCI CGAP C08"
 /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 colon adenocarcinoma, and was then primed with a Not I -
 oligo(dT) primer. Double-stranded cDNA was ligated to Eco
 RI adaptors (Pharmacia), digested with Not I and cloned
 into the Not I and Eco RI sites of the modified pT7T3
 vector. Library is normalized. Library was constructed by
 Bento Soares and M. Fatima Bonaldo."

ORIGIN
 Query Match 91.0%; Score 453; DB 9; Length 510;
 Best Local Similarity 99.6%; Pred. No. 7.7e-79;
 Matches 475; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 1 CTCAGGCGCGGCTGCGCGCTCCCGCTAGGCTCCCGGCTCAGCATCTAGTAG 60
 Db 475 CTCAGGCGCGGCTGCGCGCTCCCGCTAGGCTCCCGGCTCAGCATCTAGTAG 417
 QY 61 AGGCGCGCTTCTGGGTCAAGCTGCAACCGCGCTCTGCTCTTCTCGAAGGCGC 120
 Db 416 -GGCGCGCTTCTGGGTCAAGCTGCAACCGCGCTCTGCTCTTCTCGAAGGCGC 358
 QY 121 GTGTAGTCTCCAGTATGTTGGCCAGGCTCTTCCGACCTTCTGACCAAGCGCC 180
 Db 357 GTGTAGTCTCCAGTATGTTGGCCAGGCTCTTCCGACCTTCTGACCAAGCGCC 298
 QY 181 AAGACTGACCGCAGAGAGAGGGGCTCACTCTATCTCTCGGCGACCACTGACAAAG 240

Db 297 AAGACTGACCGCAGAGAGAGGGGCTCACTCTATCTCTCGGCGACCACTGACAAAG 238
 QY 241 CAGCGCGCTCTCCAGACTTAATATATACCACTACCTGTGAGGGGAGCCCAATCTG 300
 Db 237 CAGCGCGCTCTCCAGACTTAATATATACCACTACCTGTGAGGGGAGCCCAATCTG 178
 QY 301 GACTCTTCCCGCTTGGAGACTGACAGCGCGGAGAGAGTCCCGCAGGCTCTG 360
 Db 177 GACTCTTCCCGCTTGGAGACTGACAGCGCGGAGAGAGTCCCGCAGGCTCTG 118
 QY 361 AGGAGAGCTTCAGAGAGGCACTGACGCTGTGCGCGGAGGCTCTCGACATCCGCAAGC 420
 Db 117 AGGAGAGCTTCAGAGAGGCACTGACGCTGTGCGCGGAGGCTCTCGACATCCGCAAGC 58
 QY 421 ACCAGGAAAGTCTCTGCGGCGATCTGTAATTAACCTTTTCTTTTGTGTTT 477
 Db 57 ACCAGGAAAGTCTCTGCGGCGATCTGTAATTAACCTTTTCTTTTGTGTTT 1
 RESULT 4
 AI821606/c 537 bp mRNA linear EST 13-DEC-1999
 LOCUS nk08a11.x5 NCI CGAP C02 Homo sapiens cDNA clone IMAGE:1012892 3'
 DEFINITION similar to contains TAR1.t3 TAR1 MER22 repetitive element ;, mRNA
 sequence.
 ACCESSION AI821606
 VERSION AI821606 GI:5440685
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 537)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 Tumor Gene Index
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Unpublished (1997)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Stratagene, Inc., David B. Krizman,
 Ph.D.
 CDNA Library Arraying: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
 www-bio.llnl.gov/bdrp/image/image.html
 This read is a RESEQUENCE of a previously sequenced human clone
 Original clone citation: National Cancer Institute, Cancer Genome
 Anatomy Project (CGAP), Tumor Gene Index
 This read has been verified (found to hit its original self in the
 correct orientation)
 Insert Length: 741 Std Error: 0.00
 Seq primer: -400P from Gibco
 High quality sequence stop: 459.
 Location/Qualifiers
 1..537
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1012892"
 /tissue_type="tumor"
 /lab_host="G04R (kanamycin resistant)"
 /clone_lib="NCI CGAP C02"
 /note="Organ: colon; Vector: Bluescript SK-; Site: 1:
 EcoRI; Site: 2: XhoI; Cloned unidirectionally. Primer:
 Oligo dT. Bulk colon villous adenoma. 5' adaptor sequence:
 5' GAATTCGACGAG 3' 3' adaptor sequence: 5'
 CTCGAGTTTCTTTTCTTTTCTTTT 3' Average insert size: 1.1 kb."

ORIGIN

Query Match 89.7%; Score 446.6; DB 9; Length 537;
 Best Local Similarity 98.7%; Pred. No. 1.4e-77;
 Matches 471; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 1 CTCTAGCGTCCGCTCTGCGCCGCTCCGCTAGGCTCTCCGCGCTCAGCACTAGTACG 60
 Db CTCTAGCGTCCGCTCTGCGCCGCTCTCCGCTAGGCTCTCCGCGCTCAGCACTAGTACG 418
 QY 476 CTCTAGCGTCCGCTCTGCGCCGCTCTCCGCTAGGCTCTCCGCGCTCAGCACTAGTACG 418
 Db 61 AGCGCGGCTTTCTGGGCTCAGCTGCGCAACCGCGCTCTCTGCTCTTCTCTCGAGAGGCGC 120
 QY 417 -GCCTCGCTTCTGGGCTCAGCTGCGCAACCGCGCTCTCTGCTCTTCTCTCGAGAGGCGC 359
 Db 121 GTGTAGTCTCTCAATATGTTGGCGCCAGCGCTCTTGCACCTTCTGACCAAGAGCGC 180
 Db 358 GTGTAGTCTCTCAATATGTTGGCGCCAGCGCTCTTGCACCTTCTGACCAAGAGCGC 229
 QY 181 AAGGACTGAGCCAG 240
 Db 298 AAGGACTGAGCCAG 239
 QY 241 CAGGCGGCTCTCCGCTCTGAGCTTAAATATATCACTAATCTGTAGAGAGAGAGAGAGAG 300
 Db 238 CAGGCGGCTCTCCGCTCTGAGCTTAAATATATCACTAATCTGTAGAGAGAGAGAGAGAG 179
 QY 301 GACTCTTCCCGCTTGGGACATGCAAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
 Db 178 GACTCTTCCCGCTTGGGACATGCAAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 119
 QY 361 AAGGAGCTCCAG 420
 Db 118 AAGGAGCTCCAG 59
 QY 421 ACCAGGAAAGTCTCTCTGGGCGATCTGTAAATAAAGCTTTTCTTTTGTGTTT 477
 Db 58 ACCAGGAAAGTCTCTCTGGGCGATCTGTAAATAAAGCTTTTCTTTTGTGTTT 2

RESULT 5 446 bp mRNA linear EST 27-OCT-1999
 A1984141/c 446 bp mRNA linear EST 27-OCT-1999
 LOCUS WU21C02.x1 Soares Dieckgraebe.colon_NHCD Homo sapiens cDNA clone
 DEFINITION IMAGE:2520674.3, mRNA sequence.
 ACCESSION A1984141.
 VERSION A1984141.1 GI:5811360
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Fukuyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 TITLE Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 JOURNAL NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seg primer: 40up from Gibco.

FEATURES

source
 1..446
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2520674"
 /tissue_type="colonic mucosa from 3 patients with Crohn's
 disease"
 /lab_host="DH10B (phage-resistant)"
 /clone_1ib="Soares Dieckgraebe.colon_NHCD"
 /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; Site_1: Not 1; Site_2: Eco RI; 1st

ORIGIN

Query Match 87.3%; Score 435; DB 9; Length 446;
 Best Local Similarity 99.8%; Pred. No. 2.7e-75;
 Matches 446; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 34 GCTCTCCGCGCTTACCACTAGTACGAGCGCCCTTCTGGGTACGCTGGCAACCGGC 93
 Db 446 GCTCTCCGCGCTTACCACTAGTACGAGCGCCCTTCTGGGTACGCTGGCAACCGGC 388
 QY 94 GTCTGTGCTCTTCTCTCTGAG 153
 Db 387 GTCTGTGCTCTTCTCTCTGAG 328
 QY 154 CTTGCAACCTTCTGACCAAGAGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 213
 Db 327 CTTGCAACCTTCTGACCAAGAGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 268
 QY 214 CTATATCTCCGCGAGCCAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 273
 Db 267 CTATATCTCCGCGAGCCAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 208
 QY 274 ACTAATCTGTAG 333
 Db 207 ACTAATCTGTAG 148
 QY 334 GGAAGCATGCGCCCGCAGGCTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 393
 Db 147 GGAAGCATGCGCCCGCAGGCTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 88
 QY 394 GAGCGAGAGAGCTCGAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 453
 Db 87 GAGCGAGAGAGCTCGAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 28
 QY 454 AAACCTTTTCTTTTGTGTTTAA 480
 Db 27 AAACCTTTTCTTTTGTGTTTAA 1

RESULT 6 466 bp mRNA linear EST 12-SEP-1997
 AAS73825/c 466 bp mRNA linear EST 12-SEP-1997
 LOCUS nK08a11.s1 NCI_CGAP_C02 Homo sapiens cDNA clone IMAGE:1012892.3,
 DEFINITION mRNA sequence.
 ACCESSION AAS73825
 VERSION AAS73825.1 GI:2348340
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Fukuyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 TITLE Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 JOURNAL NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
 Emert-Buck, M.D., Ph.D.
 cDNA Library Preparation: Stratagene, Inc., David B. Kitzman,

Ph.D.
 CDNA Library Arraying: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html
 Insert Length: 741 Std Error: 0.00
 Seq primer: -40ml3 fwd. RT from Amersham
 High quality sequence stop: 456.
 Location/Qualifiers

REFERENCE
 1 (bases 1 to 434)
 AUTHORS
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL
 Unpublished (1997)
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-rc@mail.nih.gov
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 558 Std Error: 0.00
 Seq primer: -40up from Gibco.

FEATURES

Source

1. .466
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1012892"
 /issue_type="tumor"
 /lab_host="SOLR (kanamycin resistant)"
 /clone_lib="NCI CGAP C02"
 /note="Organ: colon; Vector: Bluescript SK-; Site:1;
 EcoRI; Site:2: XhoI; cloned unidirectionally. Primer:
 Oligo dt. Bulk colon villous adenoma. 5' adaptor sequence:
 5' GAATCGGCACGAG 3' 3' adaptor sequence: 5'
 CTCGAGTTTCTTTTCTTTT 3' Average insert size: 1.1 Kb."

ORIGIN

Query Match 86.6%; Score 431.4; DB 9; Length 466;
 Best Local Similarity 99.1%; Pred. No. 1,4e-74;

Matches 465; Conservative 0; Mismatches 1; Indels 3; Gaps 3;

9 TGGCGCTCTGCGCGCTCCGCTAGGCTCTCCGCGCTCACCACCTAGTACGAGCGCCG 68
 466 TGGCGCTCTGCGCG-TCGCGCTAGGCTCTCCGCGCTCACCACCTAG-ATGAGCGCGC 409
 QY 69 CTCTGTGGGTACGCTGGCAACCGGCGTCTCTGTGCTCTCTCTCGAGAGGGCCGTGTAG 128
 Db 408 CTCTGTGGGTACGCTGGCAACCGGCGTCTCTGTGCTCTCTCTCGAGAGGGCCGTGTAG 349
 QY 129 TCTCCAGTATGTTTCGCGCCAGCGCTCTTGCACACCTTTCGACCAAGGCGCAAGACTG 188
 Db 348 TCTCCAGTATGTTTCGCGCCAGCGCTCTTGCACACCTTTCGACCAAGGCGCAAGACTG 289
 QY 189 CAGCGAGAGAGAGGGGGCTCACTCTTATCTCTGCGACCACTGACACAGAGCGGC 248
 Db 288 CAGCGAGAGAGAGGGGGCTCACTCTTATCTCTGCGACCACTGACACAGAGCGGC 229
 QY 249 TCTCCAGTATGTTTCGCGCCAGCGCTCTTGCACACCTTTCGACCAAGGCGCAAGACTG 308
 Db 228 TCTCCAGTATGTTTCGCGCCAGCGCTCTTGCACACCTTTCGACCAAGGCGCAAGACTG 169
 QY 309 CCCGCGCTTGGGACATGACAGGCGGGAGACAGTGCCTCCGACAGCTGCGCCAGAGAGC 368
 Db 168 CCCGCGCTTGGGACATGACAGGCGGGAGACAGTGCCTCCGACAGCTGCGCCAGAGAGC 109
 QY 369 TCCAGGAGAGGCACTGAGCGCTGTGCGCGAGAGGCTTGGACATCCGACGAGACAGAGGA 428
 Db 108 TCCAGGAGAGGCACTGAGCGCTG-TGGCGGAGGCTTGGACATCCGACGAGAGAGGA 50
 QY 429 AAGTCTCTGCGGAGATCTGTAATAAACCTTTTCTTTTGTGTTTTT 477
 Db 49 AAGTCTCTGCGGAGATCTGTAATAAACCTTTTCTTTTGTGTTTTT 1

RESULT 7

AI991272/c

LOCUS AI991272 434 bp mRNA linear EST 09-MAR-2000
 DEFINITION w41104.x1 Soares Dieckgraebe colon_NHCD Homo sapiens cDNA clone
 IMAGE:2522647 3', mRNA sequence.

ACCESSION

AI991272

VERSION AI991272.1 GI:5838177

KEYWORDS

EST.

SOURCE

Homo sapiens

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

FEATURES

Source

1. .434
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2522647"
 /issue_type="colonic mucosa from 3 patients with Crohn's
 disease"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="Soares Dieckgraebe, colon_NHCD"
 /note="Organ: colon; Vector: pTV73D-Pac (Pharmacia) with a
 modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 TGTTCACATCGAATGAGGAGCGCGCCGCTTTTCTTTTCTTTT 3']
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pTV73 vector. Library
 went through one round of normalization. Tissue samples
 provided by Dr. Brian Dieckgraebe (Washington University,
 dieck@im.wustl.edu); colonic mucosa represents a range of
 disease involvement from moderate to severe Crohn's
 disease; samples include both perforating (fistulas) and
 non-perforating samples. Library constructed by Bento
 Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 84.9%; Score 423; DB 9; Length 434;
 Best Local Similarity 99.8%; Pred. No. 6.2e-73;

Matches 434; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

44 GCTCACCACTCAGTACGAGCGCGCTTCTGGGTACGCTGGCAACCGCGCTCTGTGCC 103
 Db 434 GCTCACCACTCAGTACG-CCGCGCTTCTGGGTACGCTGGCAACCGCGCTCTGTGCC 376
 QY 104 TCTTCTCGGAGGGCCGCTGTGAGTCTTCAGTATGTTTGGCCACGCGCTTTGGCACCC 163
 Db 375 TCTTCTCGGAGGGCCGCTGTGAGTCTTCAGTATGTTTGGCCACGCGCTTTGGCACCC 316
 QY 164 TTCTGACCAAGCGCCAGAGACTGACGCCAGAGAGAGGGGGCTCACTTATCTCTG 223
 Db 315 TTCTGACCAAGCGCCAGAGACTGACGCCAGAGAGAGGGGGCTCACTTATCTCTG 256
 QY 224 GCGACCACTGCAAGAGCGCGCTTCTCCAGACTTAAATGATACCACTTAACCTGT 283
 Db 255 GCGACCACTGCAAGAGCGCGCTTCTCCAGACTTAAATGATACCACTTAACCTGT 196
 QY 284 GAGGGGAGCCCATCTGACCTCTCCCGCGCTTGGGACATCGACGAGCGGGAGAGCATG 343
 Db 195 GAGGGGAGCCCATCTGACCTCTCCCGCGCTTGGGACATCGACGAGCGGGAGAGCATG 136
 QY 344 CCCGCGAGGCTTGGGCGAGAGAGCTCCAGAGAGGCACTGAGCGCTGTGGCGGAGGC 403
 Db 135 CCCGCGAGGCTTGGGCGAGAGAGCTCCAGAGAGGCACTGAGCGCTGTGGCGGAGGC 76
 QY 404 CTGCGACATCCGAGGCACTGAGGAGAAAGTCTCTGGGCGAGTGTAAATAAACCCTTTT 463
 Db 75 CTGCGACATCCGAGGCACTGAGGAGAAAGTCTCTGGGCGAGTGTAAATAAACCCTTTT 16
 QY 464 TTCTTTTGTGTTTTT 478
 Db 15 TTCTTTTGTGTTTTT 1

RESULT 8
AA593860/c
LOCUS mRNA 414 bp mRNA linear EST 25-SEP-1997
DEFINITION mRNA593860.1 NCI_CGAP_Col2 Homo sapiens cDNA clone IMAGE:1084349 3',
AA593860
VERSION AA593860.1 GI:2408538
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
1 (bases 1 to 414) Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Stratagene, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnhi.gov/bbrp/image/image.html
Insert Length: 1204 Std Error: 0.00
Seq primer: -40m13 fwd. Rr from Amersham
High quality sequence stop: 406.
Location/Qualifiers

FEATURES

1..414
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1084349"
/sex="mixed"
/tissue_type="colon tumor"
/lab_host="SOBR (karamycin resistant)"
/clone_lib="NCI_CGAP_Col2"
/note="Organ: colon; Vector: Bluescript SK-; Site: 1;
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Pooled colon tumors. 5' adaptor sequence: 5'
GAATCGGACGAG 3' 3' adaptor sequence: 5'
CTGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.2 kb."

ORIGIN

Query Match 83.1%; Score 414; DB 9; Length 414;
Best Local Similarity 100.0%; Pred. No. 3,7e-71;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 63 CCGCGCTTCTGGGTCACGCTGGGCAACGGGCTCTGCTCTCTCTCGAGAGGGCGCT 122
DB 414 CCGCGCTTCTGGGTCACGCTGGGCAACGGGCTCTGCTCTCTCTCGAGAGGGCGCT 355
QY 123 GGTGATCTCCAGTATGTTGGCCCGAGCGCTTTTCGACCCCTTTCGACCAAGCGCCAA 182
DB 354 GGTGATCTCCAGTATGTTGGCCCGAGCGCTTTTCGACCCCTTTCGACCAAGCGCCAA 285
QY 183 GGATTCGACGACGAGAGAGGGGGCTCATCTTTATCTCTGGGCAACCACTGCAACAGCA 242
DB 294 GGAATGACGACGAGAGAGGGGGCTCATCTTTATCTCTGGGCAACCACTGCAACAGCA 225
QY 243 GGGCGCTCTCCAGACTTAAATGTATACCACTAAGCTGTGAGGGGGAGCCCAATCTGA 302
DB 234 GGGCGCTCTCCAGACTTAAATGTATACCACTAAGCTGTGAGGGGGAGCCCAATCTGA 175
QY 303 CTCCTTCCCGGCTTGGGACATCGACAGGCGGGAGAGAGAGAGAGAGAGAGAGAGAG 362
DB 174 CTCCTTCCCGGCTTGGGACATCGACAGGCGGGAGAGAGAGAGAGAGAGAGAGAGAG 115

QY 363 GAGAGCTCCAGAGAGGCACTGACCGTCTGGCGGAGAGCCCTCGACATCCGACGAC 422
DB 114 GAGAGCTCCAGAGAGGCACTGACCGTCTGGCGGAGAGCCCTCGACATCCGACGAC 55
QY 423 CAGGAAAGTCTCTGGGGGAGATCTGAAATTAACCTTTTCTTTTCTTTT 476
DB 54 CAGGAAAGTCTCTGGGGGAGATCTGAAATTAACCTTTTCTTTTCTTTT 1

RESULT 9
A1983793/c
LOCUS mRNA 411 bp mRNA linear EST 27-OCT-1999
DEFINITION W420C09.X1 Soares_Dieckgraefe_colon_NHCD Homo sapiens cDNA clone
IMAGE:2520592 3', mRNA sequence.
A1983793
VERSION A1983793.1 GI:5811012
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
1 (bases 1 to 411) Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL

COMMENT

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnhi.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 400.
Location/Qualifiers

FEATURES

1..411
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2520592"
/tissue_type="colonic mucosa from 3 patients with Crohn's
disease"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Soares_Dieckgraefe_colon_NHCD"
/note="Organ: colon; Vector: pTZ19-Pac (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - Oligo(dt) primer [5'
TGTACCATCTGATGAGGAGCGCGGCTTTTCTTTTCTTTTCTTTT 3']
TGTACCATCTGATGAGGAGCGCGGCTTTTCTTTTCTTTTCTTTT 3']
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pTZ19 vector. Library
went through one round of normalization. Tissue samples
provided by Dr. Brian Dieckgraefe (Washington University,
dieck@im.wustl.edu); colonic mucosa represents a range of
disease involvement from moderate to severe Crohn's
disease; samples include both perforating (fistulas) and
non-perforating samples. Library constructed by Bento
Soares and W. Fatima Bonaldo."

ORIGIN

Query Match 82.2%; Score 409.4; DB 9; Length 411;
Best Local Similarity 99.8%; Pred. No. 2,9e-70;
Matches 410; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 72 CTGGGTACGCTGGCAACCGCGCTCTGTGCTCTTCTTCTGAGGGGCGGTGTAGTCT 131
DB 411 CTGGGTACGCTGGCAACCGCGCTCTGTGCTCTTCTTCTGAGGGGCGGTGTAGTCT 352
QY 132 CCAATATGTTGGGCCAGCGCTCTTTCGACCCCTTTCGACCAAGAGCCCAAGAGCTGAG 191
DB 351 CCAATATGTTGGGCCAGCGCTCTTTCGACCCCTTTCGACCAAGAGCCCAAGAGCTGAG 292
QY 192 CCAGAGAGAGGGGCTTACCTCTTATCTTCTGAGACCACTGCAACAAGAGGCGCTCT 251
DB 291 CCAGAGAGAGGGGCTTACCTCTTATCTTCTGAGACCACTGCAACAAGAGGCGCTCT 232

QY 252 CCAGACTTAAATGATACACCACTAAGCTGAGGGAGGCCAATCTGACTCTCTCC 311
Db 231 CCAGACTTAAATGATACACCACTAAGCTGAGGGAGGCCAATCTGACTCTCTCC 172
QY 312 CGCTTGGGACATCGACGCGGGAGACAGTGTCCGCCAGGCTTGAGCCAGAGAGCTTC 371
Db 171 CGCTTGGGACATCGACGCGGGAGAGAGTGTCCGCCAGGCTTGAGCCAGAGAGCTTC 112
QY 372 AGAAGGGGACATGAGCGCTGCGCGAGAGGCTCGACATCGAGGACACAGGGAAAG 431
Db 111 AGAAGGGGACATGAGCGCTGCGCGAGAGGCTCGACATCGAGGACACAGGGAAAG 52
QY 432 TCTCTGGGCGATCTGTAAATTAACCTTTTCTTTGTTTTTAAAAA 482
Db 51 TCTCTGGGCGATCTGTAAATTAACCTTTTCTTTGTTTTTAAAAA 1

RESULT 10
AI274929/c 405 bp mRNA linear EST 29-JAN-1999
LOCUS q149c11.x1 NCI_CGAP_C08 Homo sapiens cDNA clone IMAGE:1875668 3'
DEFINITION similar to contains TAR1.t3 TAR1 repetitive element ;, mRNA
sequence.
AI274929
VERSION AI274929.1 GI:3897203
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 405)
NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.lnl.gov/bbrp/image/image.html
Insert Length: 1458 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 395.
Location/Qualifiers
1..405
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1875668"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_C08"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."

ORIGIN
Query Match 81.3%; Score 405; DB 9; Length 405;
Best Local Similarity 100.0%; Pred.No. 2.le-69;
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 73 TGGGTACCGCTGCAACCGGCGTCTGTCTCTTCTCGAGGGGCGTGTAGTCTC 132

Db 405 TGGGTACCGCTGCAACCGGCGTCTGTCTCTTCTCGAGGGGCGTGTAGTCTC 346
QY 133 CAGTATGTTGGGCGGAGCGCTTGTGGACCTTCTGACCCAAAGGCCAAGATGTGAGC 192
Db 345 CAGTATGTTGGGCGGAGCGCTTGTGGACCTTCTGACCCAAAGGCCAAGATGTGAGC 286
QY 193 CAGAGAGAGGGGGCTCACTCTTATCTCTGCGAGCCCACTGACCAAGCAGGCGCTTC 252
Db 285 CAGAGAGAGGGGGCTCACTCTTATCTCTGCGAGCCCACTGACCAAGCAGGCGCTTC 226
QY 253 CCAGACTTAAATGATACACCACTAAGCTGAGGGAGGCCAATCTGACTCTCTCC 312
Db 225 CCAGACTTAAATGATACACCACTAAGCTGAGGGAGGCCAATCTGACTCTCTCC 166
QY 313 GCCTTGGGACATCGACGCGGGAGAGAGTGTCCGCCAGGCTTGAGCCAGAGAGCTCA 372
Db 165 GCCTTGGGACATCGACGCGGGAGAGAGTGTCCGCCAGGCTTGAGCCAGAGAGCTCA 106
QY 373 GGAAGGGGACATGAGCGCTGCGCGAGAGGCTCGACATCGAGGACACAGGGAAAGT 432
Db 105 GGAAGGGGACATGAGCGCTGCGCGAGAGGCTCGACATCGAGGACACAGGGAAAGT 46
QY 433 CTCTTGGGCGATCTGTAAATTAACCTTTTCTTTGTTTTT 477
Db 45 CTCTTGGGCGATCTGTAAATTAACCTTTTCTTTGTTTTT 1

RESULT 11
AI346155/c 405 bp mRNA linear EST 02-FEB-1999
LOCUS gp43f12.x1 NCI_CGAP_C08 Homo sapiens cDNA clone IMAGE:1925807 3'
DEFINITION similar to contains TAR1.t3 TAR1 repetitive element ;, mRNA
sequence.
AI346155
VERSION AI346155.1 GI:4083361
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 405)
NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.lnl.gov/bbrp/image/image.html
Insert Length: 1488 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 375.
Location/Qualifiers
1..405
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1925807"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_C08"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned

ORIGIN

into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo. "

Query Match 80.4%; Score 400.2; DB 9; Length 405;
Best Local Similarity 99.3%; Pred. No. 1.9e-68;
Matches 402; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 73 TGGGTACGCTGGCAACCGCGCTCTGTCCTCTCTCTGAGGGGCGGTGAGTCTC 132
Db 405 TGGGTACGCTGGCAACCGCGCTCTGTCCTCTCTCTGAGGGGCGGTGAGTCTC 346
QY 133 CAGTATGTTGCGCCCGCGCTCTGTCCTCTCTCTGAGGGGCGGTGAGTCTC 192
Db 345 CAGTATGTTGCGCCCGCGCTCTGTCCTCTCTCTGAGGGGCGGTGAGTCTC 286
QY 193 CAGGAGAGAGGGGGCTCACCTCTTATCTCTGCGAGCCACTGACAGAGGGGCTCTC 252
Db 285 CAGGAGAGAGGGGGCTCACCTCTTATCTCTGCGAGCCACTGACAGAGGGGCTCTC 226
QY 253 CCGACTTAAATGATACCACTAATCTGAGGGGAGCCCAATCTGAGTCTCTCC 312
Db 225 CCGACTTAAATGATACCACTAATCTGAGGGGAGCCCAATCTGAGTCTCTCC 166
QY 313 GCCTTGGGACATGCGAGGCGGGAGAGAGTGCCTGCGAGGCTGGGCGGAGAGAGTCCA 372
Db 165 GCCTTGGGACATGCGAGGCGGGAGAGAGTGCCTGCGAGGCTGGGCGGAGAGAGTCCA 106
QY 373 GGAAGGGCACTGAGCGCTGTCGGCGGAGGCTCGGAGATCCGAGGAGCGAGGAAAGT 432
Db 105 GGAAGGGCACTGAGCGCTGTCGGCGGAGGCTCGGAGATCCGAGGAGCGAGGAAAGT 46
QY 433 CTCTGGGGGCGATCTGTAAATAACCTTTTCTTTCTTTTCTTTT 477
Db 45 CTCTGGGGGCGATCAAAAATAACCTTTTCTTTTCTTTTCTTTT 1

RESULT 12
AI660493/c 538 bp mRNA linear EST 18-DEC-1999
LOCUS AI660493
DEFINITION we6702.x1 Soares, Dieckgraefe, colon_NHCD Homo sapiens cDNA clone
IMAGE:2346195.3, similar to contains TARI.c3 TARI repetitive
element; mRNA sequence.

ACCESSION AI660493
VERSION AI660493.1 GI:4764063
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 538)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 676 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 463.
Location/Qualifiers
1..538

FEATURES
Source
1..538
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2346195"
/issue_type="colonic mucosa from 3 patients with Crohn's
disease"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Soares_Dieckgraefe_colon_NHCD"

ORIGIN

Query Match 79.2%; Score 394.4; DB 9; Length 538;
Best Local Similarity 95.1%; Pred. No. 2.3e-67;
Matches 407; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 51 ACTCAGTACGAGGCGCGCTCTGAGTACGCTGGCAACCGGCTCTGCTCTCTCT 110
Db 428 ACTCAGTACGCGCTCTCTCTCTGATCCCAACCGGCTCTGCTCTCTCTCTCT 369
QY 111 CGAGAGGGGCGGTGAGTCTCCAGTATGTTGGGCGGAGGCTTCCGACCTCTGGA 170
Db 368 CGAGAGGGGCGGTGAGTCTCCAGTATGTTGGGCGGAGGCTTCCGACCTCTGGA 309
QY 171 CCAGAGCGCCAGAGACTGACAGCCAGAGAGAGGGGCTCACTTATCTCTCGAGCC 230
Db 308 CCAGAGCGCCAGAGACTGACAGCCAGAGAGAGGGGCTCACTTATCTCTCGAGCC 249
QY 231 ACTGACAGAGAGGCGCGCTCTCCAGTATGATATACCACTAATCTGAGGGG 290
Db 248 ACTGACAGAGAGGCGCGCTCTCCAGTATGATATACCACTAATCTGAGGGG 189
QY 291 ACCCAATCTGAGTCTCTCTCTCTGAGTACGAGGCGGAGAGAGTCCCGCA 350
Db 188 ACCCAATCTGAGTCTCTCTCTCTGAGTACGAGGCGGAGAGAGTCCCGCA 129
QY 351 GGCCTGGGCGCAGAGAGTCTCCAGAGAGGCACTGAGGCTGCGGCGGAGGCTCGAC 410
Db 128 GGCCTGGGCGCAGAGAGTCTCCAGAGAGGCACTGAGGCTGCGGCGGAGGCTCGAC 69
QY 411 ATCCGAGGAGCAGAGGAGAGTCTCTGAGGCGGAGTCTGTAATTAACCTTTTCTTT 470
Db 68 ATCCGAGGAGCAGAGGAGAGTCTCTGAGGCGGAGTCTGTAATTAACCTTTTCTTT 9
QY 471 GTTTTGA 478
Db 8 GTTTTGA 1

RESULT 13
CB854121/c 573 bp mRNA linear EST 22-APR-2003
LOCUS CB854121
DEFINITION UI-CF-DUI-aal-1-16-0-UI.s1 UI-CF-DUI Homo sapiens cDNA clone
IMAGE:2346195.3, similar to contains TARI.c3 TARI repetitive
element; mRNA sequence.

ACCESSION CB854121
VERSION CB854121.1 GI:30044498
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 573)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery

JOURNAL Genome Res. 6 (3), 791-806 (1996)
MEDLINE 9704447
PUBMED 8889548
COMMENT Contact: McCray, PB

/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAGTGGAGGGGCGGCGCTTTTCTTTTCTTTTCTTTT 3']
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Tissue samples
provided by Dr. Brian Dieckgraefe (Washington University,
dieckgrm.wustl.edu); colonic mucosa represents a range of
disease involvement from moderate to severe Crohn's
disease; samples include both perforating (fistulas) and
non-perforating samples. Library constructed by Bento
Soares and M. Fatima Bonaldo. "

McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@iowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com)
The following repetitive elements were found in this CDNA
Sequence: 533573, XAUT (matched complement)
Seq primer: M13 FORWARD
PolyA=No.

FEATURES

Source

Location/Qualifiers
1..573
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-DUI-aal-1-16-0-UI"
/tissue_type="Primary Lung Epithelial Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-DUI"
/note="Organ: lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site: 1: EcoR I; Site: 2: Not I;
UI-CF-DUI is a normalized CDNA library containing the
following tissue(s): Primary Lung Epithelial Cells
Library was constructed according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. First strand
CDNA synthesis was primed with an oligo-dT primer
containing a Not I site. Double stranded cDNA was ligated
to an EcoR I adaptor, digested with Not I, and cloned
directionally into pT73-Pac vector. The oligonucleotide
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I
site and the (dT)18 tail. The sequence tag for this
library is GGCCTGAGGC.
TAG_SEQ=None found"

ORIGIN

Query Match 77.8%; Score 387.4; DB 14; Length 573;
Best Local Similarity 94.9%; Pred. No. 5.3e-66;
Matches 411; Conservative 0; Mismatches 21; Indels 1; Gaps 1;
QY 51 ACTGAGTAGAGAGGCGCCCTTCTGAGTCAAGCTGCAACCGGCTCTCTT-CC 109
DB 540 ACTGACATCCCTCTTCTTTCGATCCCAACCGGCAAGGCTCTGCTCTTCCC 481
QY 110 TCGAGAGGCGCGTGTGATGTCACAGTATGTCGCCAGCGCTTTCGACCTTCGG 169
DB 480 TCGGAGGCGCGTGTGATGTCACAGTATGTCGCCAGCGCTTTCGACCTTCGG 421
QY 170 ACCAAGCGCCAGAGTGTGACGACGAGAGAGAGGCGCTCACCTTTATCTCGGCAAC 229
DB 420 ACCAAGCGCCAGAGTGTGACGACGAGAGAGAGGCGCTCACCTTTATCTCGGCAAC 361
QY 230 CACTGCAAGACGAGCGCTCTCCAGACTTAATAATGTATCAACCTAACCTGTAGAGGG- 289
DB 360 CACTGCAAGACGAGCGCTCTCCAGACTTAATAATGTATCAACCTAACCTGTAGAGGG- 301
QY 290 GACCCATCTGAGACTCTTCCCGCTTGGAGCATGCAAGGCGGAGAAAGCAATGCCCGCC 349
DB 300 GACCCATCTGAGACTCTTCCCGCTTGGAGCATGCAAGGCGGAGAAAGCAATGCCCGCC 241
QY 350 AGGCTGTGGCCAGAGAGCTCCAGAGAGGCACTAGAGGCTGTGAGCGCGAGGCTCGGA 409
DB 240 AGGCTGTGGCCAGAGAGCTCCAGAGAGGCACTAGAGGCTGTGAGCGCGAGGCTCGGA 181
QY 410 CATCCGAGGCAACAGAGAAAGTCTCTGAGGCGATCTGTAAATTAACCTTTTCTTT 469

DB 180 CATCCGAGGCAACAGAGAAAGTCTCTGAGGCGATCTGTAAATTAACCTTTTCTTT 121
QY 470 TGTTTTAAAAA 482
DB 120 TGTTTTAAAAA 108

RESULT 14
LOCUS A1281211/c
DEFINITION A1281211 381 bp mRNA linear EST 28-JAN-1999
q589e08.x1 NCI CGAP Co8 Homo sapiens CDNA clone IMAGE:1873190 3'
similar to contains TAR1.t3 TAR1 repetitive element /, mRNA
sequence.
A1281211
A1281211 GI:3919444
EST.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 381)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNM, at:
www.bio.11n1.gov/bbrp/image/image.html
Insert Length: 1422 Std Error: 0.00
Seq primer: -40up from c1bco
High quality sequence stop: 340.

FEATURES
Source
Location/Qualifiers
1..381
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1873190"
/tissue_type="adenoecarcinoma"
/lab_host="DH10B"
/clone_lib="NCI CGAP Co8"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo. "

FEATURES

Source

Query Match 75.5%; Score 381; DB 9; Length 381;
Best Local Similarity 100.0%; Pred. No. 1.1e-64;
Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 97 CTGTGCTCTTCTCGAGAGGCGCGTGTGATCTCCAGTATGTTCCGCGCAGGCTCTT 156
DB 381 CTGTGCTCTTCTCGAGAGGCGCGTGTGATCTCCAGTATGTTCCGCGCAGGCTCTT 322
QY 157 CGACCTCTTCTGAGCAAGAGCCCAAGACTGCAAGCAAGAGAGAGGCGCTCACTCTT 216
DB 321 CGACCTCTTCTGAGCAAGAGCCCAAGACTGCAAGCAAGAGAGGCGCTCACTCTT 262
QY 217 ATCTCTGAGCAACCACTGCAAGAGAGCGCGCTCTCCAGACTTAATAATGTATCAACCACT 276
DB 261 ATCTCTGAGCAACCACTGCAAGAGAGCGCGCTCTCCAGACTTAATAATGTATCAACCACT 202

ORIGIN

```

QY 277 AACCTGTAGGGGGGACCAATCTGAGACTCTTCCCGCTTGGGACATCCGACGCCGGGA 336
DB 201 AACCTGTAGGGGGGACCAATCTGAGACTCTTCCCGCTTGGGACATCCGACGCCGGGA 142
QY 337 AGCAGTGGCCCGCAGGAGCTTGGGCGAGAGCTCCAGGAGGAGGAGCTGAGGCTGTGGC 396
DB 141 AGCAGTGGCCCGCAGGAGCTTGGGCGAGAGCTCCAGGAGGAGGAGCTGAGGCTGTGGC 82
QY 397 GGGAGGCTTCGAGACATCCGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 456
DB 81 GGGAGGCTTCGAGACATCCGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 22

QY 457 CCTTTTCTTTCTTTTCTTTT 477
DB 21 CCTTTTCTTTCTTTTCTTTT 1

RESULT 15
CB305399 737 bp mRNA linear EST 04-MAR-2003
LOCUS UI-CF-EN1-aed-m-17-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone
ACCESSION CB305399
VERSION CB305399.1 GI:28845910
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 737)
AUTHORS Ronaldo, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: McCray, PB
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
CDNA Sequencing by: Dr. M. Bento Soares, University of Iowa
CDNA Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
The following repetitive elements were found in this cDNA
sequence: 29-138, >MIR#SINE/MIR (matched complement) 195-220, >ALU
223-320, >ALU
Seq primer: M13 FORWARD
POLYA=yes.

FEATURES
Source
Location/Qualifiers
1..737
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-EN1-aed-m-17-0-UI"
/tissue_type="Primary Lung Cystic Fibrosis Epithelial
Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-EN1"
/note="Organ: Lung; Vector: pTZ19-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-CF-EN1 is a normalized cDNA library containing the
following tissue(s): Primary Lung Cystic Fibrosis
Epithelial Cells. The library was constructed according to
Ronaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. First strand cDNA synthesis was primed with an

```

oligo-gt primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pTZ19-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)₁₈ tail. The sequence tag for this library is CTGCTCAGGT. TAG Tissue=Human Lung Epithelial Cell lines untreated LPS 6hr to LPS 24h TAG_Lib=UI-CF-EN1 TAG_SEQ=CTGCTCAGGT"

ORIGIN

```

Query Match 75.9%; Score 378.2; DB 14; Length 737;
Best Local Similarity 92.9%; Pred. No. 3.1e-64;
Matches 395; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 51 ACTCAGTACGAGCGCGCTTCTGGGTACGCTGGCAACCGGGTCTGTGCTCTTCT 110
DB 313 ACTCCAGATGCGCTCTCTTCTTTGTGATCCCGACCGCCAGCGGTCTGTGCTCTCT 372
QY 111 CGGAGGGGCGGNGAGAGTCTCCAGTATGTTGGGCCAGGCGCTTTCGACCTTGTGA 170
DB 373 CGGAGGGGCGGTGTGAGTCTCCAGTATGTTGGGCCAGGCGCTTTCGACCTTGTGA 432
QY 171 CCAAGGCCCAAGGACTGACGACGAGAGAGGGGGCTCACTCTTATCTTGGCGACCC 230
DB 433 CCAAGGCCCAAGGACTGACGACGAGAGAGGGGGCTCACTCTTATCTTGGCGACCC 492
QY 231 ACTGCACAGAGGCGCGCTCTCCAGACTTAATATATATACCACTTAACCTGTAGGGGG 290
DB 493 ACTGCACAGAGGCGCGCTCTCCAGACTTAATATATATACCACTTAACCTGTAGGGGG 552
QY 291 ACCCAATGTGACTCTTCCCGGCTTGGGACATCGAGGCGCGGAGAGAGTCCCGCCA 350
DB 553 ACCCAATGTGACTCTTCCCGGCTTGGGACATCGAGGCGCGGAGAGAGTCCCGCCA 612
QY 351 GGCCCTGGGCGCAGAGAGTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 410
DB 613 GGCCCTGGGCGCAGAGAGTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 672
QY 411 ATCCGACGACCAAGGAGGAGTCTCTGAGGCGATCTGTAATTAACCTTTTCTTTT 470
DB 673 ATCCGACGACCAAGGAGGAGTCTCTGAGGCGATCTGTAATTAACCTTTTCTTTT 732
QY 471 GTTTT 475
DB 733 TTTT 737

```

Search completed: February 23, 2004, 23:20:18
 Job time : 1268.96 secs